Interactive Unawareness Revisited*

Joseph Y. Halpern
Computer Science Department
Cornell University, U.S.A.
e-mail: halpern@cs.cornell.edu

Leandro Chaves Rêgo
School of Electrical and Computer Engineering
Cornell University, U.S.A.
e-mail: lcr26@cornell.edu

*A preliminary version of this paper was presented at the Tenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK05).

Abstract

We analyze a model of interactive unawareness introduced by Heifetz, Meier and Schipper (HMS). We consider two axiomatizations for their model, which capture different notions of validity. These axiomatizations allow us to compare the HMS approach to both the standard (S5) epistemic logic and two other approaches to unawareness: that of Fagin and Halpern and that of Modica and Rustichini. We show that the differences between the HMS approach and the others are mainly due to the notion of validity used and the fact that the HMS is based on a 3-valued propositional logic.

Keywords: Interactive epistemology, awareness, modal logic, axiomatization.

1 Introduction

Reasoning about knowledge has played a significant role in work in philosophy, economics, and distributed computing. Most of that work has used standard Kripke structures to model knowledge, where an agent knows a fact φ if φ is true in all the worlds that the agent considers possible. While this approach has proved useful for many applications, it suffers from a serious shortcoming, known as the *logical omniscience* problem (first observed and named by Hintikka [1962]): agents know all tautologies and know all the logical consequences of their knowledge. This seems inappropriate for resource-bounded agents and agents who are unaware of various concepts (and thus do not know logical tautologies involving those concepts). To take just one simple example, a novice investor may not be aware of the notion of the price-earnings ratio, although that may be relevant to the decision of buying a stock.

There has been a great deal of work on the logical omniscience problem (see [Fagin, Halpern, Moses, and Vardi 1995] for an overview). Of most relevance to this paper are approaches that have focused on (lack of) awareness. Fagin and Halpern [1988] (FH from now on) were the first to deal with lack of model omniscience explicitly in terms of awareness. They did so by introducing an explicit awareness operator. Since then, there has been a stream of papers on the topic in the economics literature (see, for example, [Modica and Rustichini 1994; Modica and Rustichini 1999; Dekel, Lipman, and Rustichini 1998]). In these papers, awareness is defined in terms of knowledge: an agent is aware of p if he either knows p or knows that he does not know p. All of them focused on the single-agent case. Recently, Heifetz, Meier, and Schipper [2003] (HMS from now on) have provided a multi-agent model for unawareness. In this paper, we consider how the HMS model compares to other work.

A key feature of the HMS approach (also present in the work of Modica and Rustichini [1999]—MR from now on) is that with each world or state is associated a (propositional) language. Intuitively, this is the language of concepts defined at that world. Agents may not be aware of all these concepts. The way that is modeled is that in all the states an agent considers possible at a state s, fewer concepts may be defined than are defined at state s. Because a proposition p may be undefined at a given state s, the underlying logic in HMS is best viewed as a 3-valued logic: a proposition p may be true, false, or undefined at a given state.

We consider two sound and complete axiomatizations for the HMS model, that differ with respect to the language used and the notion of validity. One axiomatization captures *weak validity*: a formula is weakly valid if it is never false (although it may be undefined). In the single-agent case, this axiomatization is identical to that given by MR. However, in the MR model, validity is taken with respect to "objective" state, where all formulas are defined. As shown by Halpern [2001], this axiomatization is also sound and complete in the single-agent case with respect to a special case of FH's awareness structures; we extend Halpern's result to the multi-agent case. The other axiomatization of the HMS model captures *(strong) validity*: a formula is (strongly) valid if it is always true. If we add an axiom saying that there is no third value to this axiom system, then we just get the standard axiom system for S5. This shows that, when it comes to strong validity, the only difference between the HMS models and standard epistemic models is the third truth value.

The rest of this paper is organized as follows. In Section 2, we review the basic S5 model, the FH model, the MR model, and the HMS model. In Section 3, we compare the HMS approach and the FH approach, both semantically and axiomatically, much as Halpern [2001] compares the MR and FH approaches. We show that weak validity in HMS structures corresponds in a precise sense to validity in awareness structures. In Section 4, we extend the HMS language by adding a nonstandard implication operator. Doing so allows us to provide an axiomatization for strong validity. We conclude in Section 5. Further discussion of the original HMS framework and an axiomatization of strong validity in the purely propositional case can be found in the appendix.

2 Background

We briefly review the standard epistemic logic and the approaches of FH, MR, and HMS here.

2.1 Standard epistemic logic

The syntax of standard epistemic logic is straightforward. Given a set $\{1, \ldots, n\}$ of agents, formulas are formed by starting with a set $\Phi = \{p, q, \ldots\}$ of primitive propositions as well as a special formula $\top$ (which is always true), and then closing off under conjunction ($\wedge$), negation ($\neg$) and the modal operators K_i, $i = 1, \ldots, n$. Call the resulting language $\mathcal{L}_n^K(\Phi)$.[1] As usual, we define $\varphi \vee \psi$ and $\varphi \Rightarrow \psi$ as abbreviations of $\neg(\neg\varphi \wedge \neg\psi)$ and $\neg\varphi \vee \psi$, respectively.

The standard approach to giving semantics to $\mathcal{L}_n^K(\Phi)$ uses Kripke structures. A *Kripke structure for n agents (over Φ)* is a tuple $M = (\Sigma, \pi, \mathcal{K}_1, \ldots, \mathcal{K}_n)$, where Σ is a set of states, $\pi : \Sigma \times \Phi \rightarrow \{0, 1\}$ is an interpretation, which associates with each primitive propositions its truth value at each state in Σ, $\mathcal{K}_i : \Sigma \rightarrow 2^{\Sigma}$ is a *possibility correspondence* for agent i. Intuitively, if $t \in \mathcal{K}_i(s)$, then agent i considers state t possible at state s. $\mathcal{K}_i$ is *reflexive* if for all $s \in \Sigma$, $s \in \mathcal{K}_i(s)$; it is *transitive* if for all $s, t \in \Sigma$, if $t \in \mathcal{K}_i(s)$ then $\mathcal{K}_i(t) \subseteq \mathcal{K}_i(s)$;it is *Euclidean* if for all $s, t \in \Sigma$, if $t \in \mathcal{K}_i(s)$ then $\mathcal{K}_i(t) \supseteq \mathcal{K}_i(s)$.[2] A Kripke structure is reflexive (resp., reflexive and transitive; partitional) if the possibility correspondences $\mathcal{K}_i$ are reflexive (resp., reflexive and transitive; reflexive, Euclidean, and transitive). Let $\mathcal{M}_n(\Phi)$ denote the class of all Kripke structures for n agents over Φ, with no restrictions on the $\mathcal{K}_i$ correspondences. We use the superscripts r, e, and t to indicate that the $\mathcal{K}_i$ correspondences are restricted to being reflexive, Euclidean,

[1] In MR, only the single-agent case is considered. We consider multi-agent epistemic logic here to allow the generalization to HMS. In many cases, $\top$ is defined in terms of other formulas, e.g., as $\neg(p \wedge \neg p)$. We take it to be primitive here for convenience.

[2] It is more standard in the philosophy literature to take $\mathcal{K}_i$ to be a binary relation. The two approaches are equivalent, since if $\mathcal{K}_i'$ is a binary relation, we can define a possibility correspondence $\mathcal{K}_i$ by taking $t \in \mathcal{K}_i(s)$ iff $(s, t) \in \mathcal{K}_i'$. We can similarly define a binary relation given a possibility correspondence. Given this equivalence, it is easy to see that the notions of a possibility correspondence being reflexive, transitive, or Euclidean are equivalent to the corresponding notion for binary relations. We remark that while we use the standard terminology in the logic and philosophy here, other terminology has occasionally been used in the economics literature. For example, Geanakoklos [1989] calls a reflexive correspondence *nondeluded* and takes a transitive correspondence to be one where the agent *knows what he knows*. There is a deep connection between these properties of the possibility correspondence and notions such as *positive introspection*; we discuss that shortly.

and transitive, respectively. Thus, for example, $\mathcal{M}_n^{rt}(\Phi)$ is the class of all reflexive and transitive Kripke structures for n agents.

We write $(M, s) \models \varphi$ if φ is true at state s in the Kripke structure M. The truth relation is defined inductively as follows:

$$(M, s) \models p, \text{ for } p \in \Phi, \text{ if } \pi(s, p) = 1$$

$$(M, s) \models \neg\varphi \text{ if } (M, s) \not\models \varphi$$

$$(M, s) \models \varphi \wedge \psi \text{ if } (M, s) \models \varphi \text{ and } (M, s) \models \psi$$

$$(M, s) \models K_i\varphi \text{ if } (M, s') \models \varphi \text{ for all } s' \in \mathcal{K}_i(s).$$

A formula φ is said to be *valid* in Kripke structure M if $(M, s) \models \varphi$ for all $s \in \Sigma$. A formula φ is valid in a class $\mathcal{N}$ of Kripke structures, denoted $\mathcal{N} \models \varphi$, if it is valid for all Kripke structures in $\mathcal{N}$.

An *axiom system* AX consists of a collection of *axioms* and *inference rules*. An axiom is a formula, and an inference rule has the form "from $\varphi_1, \ldots, \varphi_k$ infer ψ," where $\varphi_1, \ldots, \varphi_k, \psi$ are formulas. A formula φ is *provable* in AX, denoted AX $\vdash \varphi$, if there is a sequence of formulas such that the last one is φ, and each one is either an axiom or follows from previous formulas in the sequence by an application of an inference rule. An axiom system AX is said to be *sound* for a language $\mathcal{L}$ with respect to a class $\mathcal{N}$ of structures if every formula in $\mathcal{L}$ provable in AX is valid with respect to $\mathcal{N}$. The system AX is *complete* for $\mathcal{L}$ with respect to $\mathcal{N}$ if every formula in $\mathcal{L}$ that is valid with respect to $\mathcal{N}$ is provable in AX.

Consider the following set of well-known axioms and inference rules:

Prop. All substitution instances of valid formulas of propositional logic.

K. $(K_i\varphi \wedge K_i(\varphi \Rightarrow \psi)) \Rightarrow K_i\psi$.

T. $K_i\varphi \Rightarrow \varphi$.

4. $K_i\varphi \Rightarrow K_iK_i\varphi$.

5. $\neg K_i\varphi \Rightarrow K_i\neg K_i\varphi$.

MP. From φ and $\varphi \Rightarrow \psi$ infer ψ (modus ponens).

Gen. From φ infer $K_i\varphi$.

These axioms go by various names in the economics, philosophy, and logic literature.[3] For example, the axiom T has been called the *Knowledge Axiom* [Osborne and Rubinstein 1994]; axiom 4 has been called the *Positive Introspection Axiom* [Aumann 1999; Fagin, Halpern, Moses, and Vardi 1995] and the *Axiom of Transparency* [Osborne and Rubinstein 1994]; and axiom 5 has been called the *Negative Introspection Axiom* [Aumann 1999; Fagin, Halpern, Moses, and Vardi 1995] and the *Axiom of Wisdom* [Osborne and Rubinstein 1994]. It is well known that the axioms T, 4, and 5 correspond to the requirements that the $\mathcal{K}_i$ correspondences are reflexive, transitive, and Euclidean, respectively. Let $\mathbf{K}_n$ be the axiom system consisting of the axioms Prop, K and rules MP, and Gen, and let $\mathbf{S5}_n$ be the system consisting of all the axioms and inference rules above. The following result is well known (see, for example, [Chellas 1980; Fagin, Halpern, Moses, and Vardi 1995] for proofs).

Theorem 2.1: *Let $\mathcal{C}$ be a (possibly empty) subset of $\{\mathrm{T}, 4, 5\}$ and let C be the corresponding subset of $\{r, t, e\}$. Then $\mathbf{K}_n \cup \mathcal{C}$ is a sound and complete axiomatization of the language $\mathcal{L}_n^K(\Phi)$ with respect to $\mathcal{M}_n^C(\Phi)$.*

In particular, this shows that $\mathbf{S5}_n$ characterizes partitional models, where the possibility correspondences are reflexive, transitive, and Euclidean.

[3] In the economics literature, they are typically applied to semantic knowledge operators, which are functions from sets to sets, rather than to syntactic knowledge operators, but they have essentially the same form.

2.2 The FH model

The Logic of General Awareness model of Fagin and Halpern [1988] introduces a syntactic notion of awareness. This is reflected in the language by adding a new modal operator A_i for each agent i. The intended interpretation of $A_i\varphi$ is "i is aware of φ". The power of this approach comes from the flexibility of the notion of awareness. For example, "agent i is aware of φ" may be interpreted as "agent i is familiar with all primitive propositions in φ" or as "agent i can compute the truth value of φ in time t".

Having awareness in the language allows us to distinguish two notions of knowledge: implicit knowledge and explicit knowledge. Implicit knowledge, denoted with K_i, is defined as truth in all worlds the agent considers possible, as usual. Explicit knowledge, denoted with X_i, is defined as the conjunction of implicit knowledge and awareness. Let $\mathcal{L}_n^{K,X,A}(\Phi)$ be the language extending $\mathcal{L}_n^K(\Phi)$ by closing off under the operators A_i and X_i, for $i = 1, \ldots, n$. Let $\mathcal{L}_n^{X,A}(\Phi)$ (resp. $\mathcal{L}_n^X(\Phi)$) be the sublanguage of $\mathcal{L}_n^{K,X,A}(\Phi)$ where the formulas do not mention $K_1, \ldots, K_n$ (resp., $K_1, \ldots, K_n$ and $A_1, \ldots A_n$).

An *awareness structure for n agents over Φ* is a tuple $M = (\Sigma, \pi, \mathcal{K}_1, ..., \mathcal{K}_n, \mathcal{A}_1, ..., \mathcal{A}_n)$, where $(\Sigma, \pi, \mathcal{K}_1, ..., \mathcal{K}_n)$ is a Kripke structure and $\mathcal{A}_i$ is a function associating a set of formulas for each state, for $i = 1, ..., n$. Intuitively, $\mathcal{A}_i(s)$ is the set of formulas that agent i is aware of at state s. The set of formulas the agent is aware of can be arbitrary. Depending on the interpretation of awareness one has in mind, certain restrictions on $\mathcal{A}_i$ may apply. There are two restrictions that are of particular interest here:

- Awareness is *generated by primitive propositions* if, for all agents i, $\varphi \in \mathcal{A}_i(s)$ iff all the primitive propositions that appear in φ are in $\mathcal{A}_i(s) \cap \Phi$. That is, an agent is aware of φ iff she is aware of all the primitive propositions that appear in φ.

- *Agents know what they are aware of* if, for all agents i, $t \in \mathcal{K}_i(s)$ implies that $\mathcal{A}_i(s) = \mathcal{A}_i(t)$.

Following Halpern [2001], we say that awareness structure is *propositionally determined* if awareness is generated by primitive propositions and agents know what they are aware of.

The semantics for awareness structures extends the semantics defined for standard Kripke structures by adding two clauses defining A_i and X_i:

$$(M, s) \models A_i\varphi \text{ if } \varphi \in \mathcal{A}_i(s)$$

$$(M, s) \models X_i\varphi \text{ if } (M, s) \models A_i\varphi \text{ and } (M, s) \models K_i\varphi.$$

FH provide a complete axiomatization for the logic of awareness; we omit the details here.

2.3 The MR model

We follow Halpern's [2001] presentation of MR here; it is easily seen to be equivalent to that in [Modica and Rustichini 1999].

Since MR consider only the single-agent case, they use the language $\mathcal{L}_1^K(\Phi)$. A *generalized standard model* (GSM) over Φ has the form $M = (S, \Sigma, \pi, \mathcal{K}, \rho)$, where

- S and Σ are disjoint sets of states; moreover, $\Sigma = \cup_{\Psi \subseteq \Phi} S_\Psi$, where the sets S_Ψ are disjoint. Intuitively, the states in S describe the objective situation, while the states in Σ describe the agent's subjective view of the objective situation, limited to the vocabulary that the agent is aware of.

- $\pi : S \times \Phi \Rightarrow \{0, 1\}$ is an interpretation.

- $\mathcal{K} : S \to 2^\Sigma$ is a *generalized possibility correspondence*.

- ρ is a projection from S onto Σ such that (1) if $\rho(s) = \rho(t) \in S_\Psi$ then (a) s and t agree on the truth values of all primitive propositions in Ψ, that is, $\pi(s, p) = \pi(t, p)$ for all $p \in \Psi$ and (b) $\mathcal{K}(s) = \mathcal{K}(t)$ and (2) if $\rho(s) \in S_\Psi$, then $\mathcal{K}(s) \subseteq S_\Psi$. Intuitively, $\rho(s)$ is the agent's subjective state in objective state s.

We can extend $\mathcal{K}$ to a map (also denoted $\mathcal{K}$ for convenience) defined on $S \cup \Sigma$ in the following way: if $s' \in \Sigma$ and $\rho(s) = s'$, define $\mathcal{K}(s') = \mathcal{K}(s)$. Condition 1(b) on ρ guarantees that this extension is well defined. A GSM is reflexive (resp., reflexive and transitive; partitional) if $\mathcal{K}$ restricted to Σ is reflexive (resp., reflexive and transitive; reflexive, Euclidean and transitive). Similarly, we can extend π to a function (also denoted π) defined on $S \cup \Sigma$: if $s' \in S_\Psi$, $p \in \Psi$ and $\rho(s) = s'$, define $\pi(s', p) = \pi(s, p)$; and if $s' \in S_\Psi$ and $p \notin \Psi$, define $\pi(s', p) = 1/2$.

With these extensions of $\mathcal{K}$ and π, the semantics for formulas in GSMs is identical to that in standard Kripke structures except for the case of negation, which is defined as follows:

if $s \in S$, then $(M, s) \models \neg\varphi$ iff $(M, s) \not\models \varphi$

if $s \in S_\Psi$, then $(M, s) \models \neg\varphi$ iff $(M, s) \not\models \varphi$ and $\varphi \in \mathcal{L}_1^K(\Psi)$.

Note that for states in the "objective" state space S, the logic is 2-valued; and every formula is either true or false (where we take a formula φ to be false at a state s if $\neg\varphi$ is true at s). On the other hand, for states in the "subjective" state space Σ the logic is 3-valued. A formula may be neither true nor false. It is easy to check that if $s \in S_\Psi$, then every formula in $\mathcal{L}_1^K(\Psi)$ is either true or false at s, while formulas not in $\mathcal{L}_1^K(\Psi)$ are neither true nor false. Intuitively, an agent can assign truth values only to formulas involving concepts he is aware of; at states in S_Ψ, the agent is aware only of concepts expressed in the language $\mathcal{L}_1^K(\Psi)$.

The intuition behind MR's notion of awareness is that an agent is unaware of φ if he does not know φ, does not know he does not know it, and so on. Thus, an agent is aware of φ if he either knows φ or knows he does not know φ, or knows that he does not know that he does not know φ, or .... MR show that under appropriate assumptions, this infinite disjunction is equivalent to the first two disjuncts, so they define $A\varphi$ to be an abbreviation of $K\varphi \vee K\neg K\varphi$.

Rather than considering validity, MR consider what we call here *objective validity*: truth in all objective states (that is, the states in S). Note that all classical (2-valued) propositional tautologies are objectively valid in the MR setting. MR provide a system $\mathcal{U}$ that is a sound and complete axiomatization for objective validity with respect to partitional GSM structures. The system $\mathcal{U}$ consists of the axioms Prop, T, and 4, the inference rule MP, and the following additional axioms and inference rules:

M. $K(\varphi \wedge \psi) \Rightarrow K\varphi \wedge K\psi$.

C. $K\varphi \wedge K\psi \Rightarrow K(\varphi \wedge \psi)$.

A. $A\varphi \Leftrightarrow A\neg\varphi$.

AM. $A(\varphi \wedge \psi) \Rightarrow A\varphi \wedge A\psi$.

N. $K\top$.

RE$_{sa}$. From $\varphi \Leftrightarrow \psi$ infer $K\varphi \Leftrightarrow K\psi$, where φ and ψ contain exactly the same primitive propositions.

Theorem 2.2: [Modica and Rustichini 1999] *$\mathcal{U}$ is a complete and sound axiomatization of objective validity for the language $\mathcal{L}_1^K(\Phi)$ with respect to partitional GSMs over Φ.*

2.4 The HMS model

HMS define their approach semantically, without giving a logic. We discuss their semantic approach in the appendix. To facilitate comparison of HMS to the other approaches we have considered, we define an appropriate logic. (In recent work done independently of ours [Heifetz, Meier, and Schipper 2005], HMS also consider a logic based on their approach, whose syntax and semantics is essentially identical to that described here.)

Given a set Φ of primitive propositions, consider again the language $\mathcal{L}_n^K(\Phi)$. An *HMS structure for n agents* (over Φ) is a tuple $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\})$, where (as in MR), $\Sigma = \cup_{\Psi \subseteq \Phi} S_\Psi$ is a set of states, $\pi : \Sigma \times \Phi \rightarrow \{0, 1, 1/2\}$ is an interpretation such that for $s \in S_\Psi$, $\pi(s, p) \neq 1/2$ iff $p \in \Psi$ (intuitively, all primitive propositions in Ψ are defined at states of S_Ψ), and $\rho_{\Psi',\Psi}$ maps $S_{\Psi'}$ onto S_Ψ. Intuitively, $\rho_{\Psi',\Psi}(s)$ is a description of the state $s \in S_{\Psi'}$ in the less expressive vocabulary of S_Ψ. Moreover, if $\Psi_1 \subseteq \Psi_2 \subseteq \Psi_3$, then $\rho_{\Psi_3,\Psi_2} \circ \rho_{\Psi_2,\Psi_1} = \rho_{\Psi_3,\Psi_1}$. Note that although both MR and HMS have projection functions, they have slightly different intuitions behind them. For MR, $\rho(s)$ is the subjective state (i.e., the way the world looks to the agent) when the actual objective state is s. For HMS, there is no objective state; $\rho_{\Psi',\Psi}(s)$ is the description of s in the less expressive vocabulary of S_Ψ. For $B \subseteq S_{\Psi_2}$, let $\rho_{\Psi_2,\Psi_1}(B) = \{\rho_{\Psi_2,\Psi_1}(s) : s \in B\}$. Finally, the $\models$ relation in HMS structures is defined for formulas in $\mathcal{L}_n^K(\Phi)$ in exactly the same way as it is in subjective states of MR structures. Moreover, like MR, $A_i\varphi$ is defined as an abbreviation of $K_i\varphi \vee K_i\neg K_i\varphi$.

Note that the definition of $\models$ does not use the functions $\rho_{\Psi,\Psi'}$. These functions are used only to impose some coherence conditions on HMS structures. To describe these conditions, we need a definition. Given $B \subseteq S_\Psi$, let $B^\uparrow = \cup_{\Psi' \supseteq \Psi} \rho^{-1}_{\Psi',\Psi}(B)$. Thus, $B^\uparrow$ consists of the states which project to a state in B.

1. Confinedness: If $s \in S_\Psi$ then $\mathcal{K}_i(s) \subseteq S_{\Psi'}$ for some $\Psi' \subseteq \Psi$.

2. Generalized reflexivity: $s \in \mathcal{K}_i(s)^\uparrow$ for all $s \in \Sigma$.

3. Stationarity: $s' \in \mathcal{K}_i(s)$ implies

   (a) $\mathcal{K}_i(s') \subseteq \mathcal{K}_i(s)$;

   (b) $\mathcal{K}_i(s') \supseteq \mathcal{K}_i(s)$.

4. Projections preserve knowledge: If $\Psi_1 \subseteq \Psi_2 \subseteq \Psi_3$, $s \in S_{\Psi_3}$, and $\mathcal{K}_i(s) \subseteq S_{\Psi_2}$, then $\rho_{\Psi_2,\Psi_1}(\mathcal{K}_i(s)) = \mathcal{K}_i(\rho_{\Psi_3,\Psi_1}(s))$.

5. Projections preserve ignorance: If $s \in S_{\Psi'}$ and $\Psi \subseteq \Psi'$ then $(\mathcal{K}_i(s))^\uparrow \subseteq (\mathcal{K}_i(\rho_{\Psi',\Psi}(s)))^\uparrow$.[4]

We remark that HMS combined parts (a) and (b) of stationarity into one statement (saying $\mathcal{K}_i(s) = \mathcal{K}_i(s')$). We split the condition in this way to make it easier to capture axiomatically. Roughly speaking,

[4]HMS explicitly assume that $\mathcal{K}_i(s) \neq \emptyset$ for all $s \in \Sigma$, but since this follows from generalized reflexivity we do not assume it explicitly. HMS also mention one other property, which they call *projections preserve awareness*, but, as HMS observe, it follows from the assumption that projections preserve knowledge, so we do not consider it here.

generalized reflexivity, part (a) of stationarity, and part (b) of stationarity are analogues of the assumptions in standard epistemic structures that the possibility correspondences are reflexive, transitive, and Euclidean, respectively. The remaining assumptions can be viewed as coherence conditions. See [Heifetz, Meier, and Schipper 2003] for further discussion of these conditions.

If C is a subset of $\{r, t, e\}$, let $\mathcal{H}_n^C(\Phi)$ denote the class of HMS structures over Φ satisfying confinedness, projections preserve knowledge, projections preserve ignorance, and the subset of generalized reflexivity, part (a) of stationarity, and part (b) of stationarity corresponding to C. Thus, for example, $\mathcal{H}_n^{rt}(\Phi)$ is the class of HMS structures for n agents over Φ that satisfy confinedness, projections preserve knowledge, projections preserve ignorance, generalized reflexivity, and part (a) of stationarity.

HMS consider only "partitional" HMS structures, that is, structures in $\mathcal{H}_n^{rte}(\Phi)$. However, we can get more insight into HMS structures by allowing the greater generality of considering non-partitional structures. In particular, by considering non-partitional structures, we can make precise the sense in which generalized reflexivity and parts (a) and (b) of stationarity are analogues of the assumptions that possiblity correspondences are reflexive, transitive, and Euclidean. Moreover, non-partitional structures are of interest in their own right. Philosophers have long argued that negative introspection is an inappropriate axiom for knowledge (see, for example, [Hintikka 1962]). In the economics literature, the consequences of dropping properties such as negative introspection (axiom 5) and the knowledge axiom (T) on speculation and Aumann's [1976] results about the impossibility of agreeing to disagree have been considered by Bacharach [1985], Geanakoplos [1989], and Samet [1987], among others. All this work has been done in the "standard" setting (essentially, in Kripke structures where all agents are aware of all formulas), so the results cannot just be lifted to HMS structures. However, by considering non-partitional HMS structures, we can separate the effects of awareness from the effects of the properties of knowledge. This lets us distinguish arguably "intrinsic" properties of the approach from ones that depend on these additional properties. For example, an intrinsic property of Kripke structures is that agents know all tautologies (this is basically a consequence of the inference rule Gen) and that agents know the logical consequences of their knowledge (this is the content of the axiom K). These properties hold no matter what assumptions we make about the correspondence relation. Similarly, considering non-partitional HSM structures allows us to consider the "intrinsic" properties of the HSM approach.

3 A Comparison of the Approaches

There is a difference in the philosophy behind HMS structures and awareness structures. HMS view states as *subjective*—in the agents' heads. In awareness structures, states are objective. In a given state, a formula p is true, even though no agent is aware of p. This is fairly standard in modal logic. For example, in reasoning about belief, we may want a state where p is true even though no agent believes p is true. However, there is an added subtlety that arises in dealing with awareness. What does it mean for an agent i to consider a state t possible where p is true when i is not aware of p? To make sense of this, if $t \in \mathcal{K}_i(s)$, then we implicitly assume that i's "understanding" of t is limited to the formulas that i is aware of at s (which are also the formulas that i is aware of at t, under the assumption that agents know what they are aware of).

Despite this difference in philosophy, there are quite close connections between the MR, HMS, and FH approaches. As a first step to comparing the approaches, we recall a result proved by Halpern.

Lemma 3.1: [Halpern 2001, Lemma 2.1] *If M is a partitional awareness structure where awareness is generated by primitive propositions, then*

$$M \models A_i\varphi \Leftrightarrow (X_i\varphi \vee (\neg X_i\varphi \wedge X_i\neg X_i\varphi)).$$

Halpern proves this lemma only for the single-agent case, but the proof goes through without change for the multi-agent case. Note that this equivalence does not hold in general in non-partitional structures. Thus, if we restrict attention to partitional awareness structures where awareness is generated by primitive propositions, we can define awareness just as MR and HMS do. Halpern [2001, Theorem 4.1] proves an even stronger connection between the semantics of FH and MR, essentially showing that partitional GSMs are in a sense equivalent to propositionally determined awareness structures. We prove a generalization of this result here.

If C is a subset of $\{r, t, e\}$, let $\mathcal{N}_n^{C,pd}(\Phi)$ and $\mathcal{N}_n^{C,pg}$ denote the set of propositionally determined awareness structures over Φ and the set of awareness structures over Φ where awareness is propositionally generated, respectively, whose $\mathcal{K}_i$ relations satisfy the conditions in C. Given a formula $\varphi \in \mathcal{L}_n^K(\Phi)$, let $\varphi_X \in \mathcal{L}_n^X(\Phi)$ be the formula that results by replacing all occurrences of K_i in φ by X_i. Finally, let Φ_φ be the set of primitive propositions appearing in φ.

Theorem 3.2: *Let C be a subset of $\{r, t, e\}$.*

(a) If $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\}) \in \mathcal{H}_n^C(\Phi)$, then there exists an awareness structure $M' = (\Sigma, \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi', \mathcal{A}_1, \ldots, \mathcal{A}_n) \in \mathcal{N}_n^{C,pg}(\Phi)$ such that, for all $\varphi \in \mathcal{L}_n^K(\Phi)$, if $s \in S_\Psi$ and $\Phi_\varphi \subseteq \Psi$, then $(M, s) \models \varphi$ iff $(M', s) \models \varphi_X$. Moreover, if $C \cap \{t, e\} \neq \emptyset$, then we can take $M' \in \mathcal{N}_n^{C,pd}$.

(b) If $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \mathcal{A}_1, \ldots, \mathcal{A}_n) \in \mathcal{N}_n^{C,pd}(\Phi)$, then there exists an HMS structure $M' = (\Sigma', \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi', \{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\}) \in \mathcal{H}_n^C(\Phi)$ such that $\Sigma' = \Sigma \times 2^\Phi$, $S_\Psi = \Sigma \times \{\Psi\}$ for all $\Psi \subseteq \Phi$, and, for all $\varphi \in \mathcal{L}_n^K(\Phi)$, if $\Phi_\varphi \subseteq \Psi$, then $(M, s) \models \varphi_X$ iff $(M', (s, \Psi)) \models \varphi$. If $\{t, e\} \cap C = \emptyset$, then the result holds even if $M \in (\mathcal{N}_n^{C,pg}(\Phi) - \mathcal{N}_n^{C,pd}(\Phi))$.

It follows immediately from Halpern's analogue of Theorem 3.2 [Halpern 2001, Theorem 4.1] that φ is objectively valid in GSMs iff φ_X is valid in propositionally determined partitional awareness structures. Thus, objective validity in GSMs and validity in propositionally determined partitional awareness structures are characterized by the same set of axioms.

We would like to get a similar result here. However, if we define validity in the usual way—that is, φ is valid iff $(M, s) \models \varphi$ for all states s and all HMS structures M—then it is easy to see that there are no (non-trivial) valid HMS formulas. Since the HMS logic is three-valued, besides what we will call *strong validity* (truth in all states), we can consider another standard notion of validity. A formula is *weakly valid* iff it is not false at any state in any HMS structure (that is, it is either true or undefined at every state in every HMS structure). Put another way, φ is weakly valid if, at all states where φ is defined, φ is true.

Corollary 3.3: *If $C \subseteq \{r, t, e\}$ then*

(a) if $C \cap \{t, e\} = \emptyset$, then φ is weakly valid in $\mathcal{H}_n^C(\Phi)$ iff φ_X is valid in $\mathcal{N}_n^{C,pg}(\Phi)$;

(b) if $C \cap \{t, e\} \neq \emptyset$, then φ is weakly valid in $\mathcal{H}_n^C(\Phi)$ iff φ_X is valid in $\mathcal{N}_n^{C,pd}(\Phi)$.

Halpern [2001] provides a sound and complete axiomatizations for the language $\mathcal{L}_1^{X,A}(\Phi)$ with respect to $\mathcal{N}^{C,pd}(\Phi)$, where C is either $\emptyset$, $\{r\}$, $\{r, t\}$ and $\{r, e, t\}$. It is straightforward to extend his techniques to other subsets of $\{r, e, t\}$ and to arbitrary numbers of agents. However, these axioms involve combinations of X_i and A_i; for example, all the systems have an axiom of the form $X\varphi \wedge X(\varphi \Rightarrow \psi) \wedge A\psi \Rightarrow X\psi$. There seems to be no obvious axiomatization for $\mathcal{L}_n^X(\Phi)$ that just involves axioms in the language $\mathcal{L}_n^X(\Phi)$ except for the special case of partitional awareness structures, where A_i is definable in terms of X_i (see Lemma 3.1), although this may simply be due to the fact that there are no interesting axioms for this language.

Let $\mathrm{S5}_n^X$ be the n-agent version of the axiom system $\mathrm{S5}_X$ that Halpern proves is sound and complete for $\mathcal{L}_n^X(\Phi)$ with respect to $\mathcal{N}_n^{ret,pd}(\Phi)$ (so that, for example, the axiom $X\varphi \wedge X(\varphi \Rightarrow \psi) \wedge A\psi \Rightarrow X\psi$ becomes $X_i\varphi \wedge X_i(\varphi \Rightarrow \psi) \wedge A_i\psi \Rightarrow X_i\psi$, where now we view $A_i\varphi$ as an abbreviation for $X_i\varphi \vee X_i\neg X_i\varphi$). Let $\mathrm{S5}_n^K$ be the result of replacing all occurrences of X_i in formulas in $\mathrm{S5}_n^X$ by K_i. Similarly, let $\mathcal{U}_n$ be the n-agent version of the axiom system $\mathcal{U}$ together with the axiom $A_iK_j\varphi \Leftrightarrow A_i\varphi$,[5] and let $\mathcal{U}_n^X$ be the result of replacing all instances of K_i in the axioms of $\mathcal{U}_n$ by X_i. HMS have shown that there is a sense in which a variant of $\mathcal{U}_n$ (which is easily seen to be equivalent to $\mathcal{U}_n$) is a sound and complete axiomatization for HMS

[5]The single-agent version of this axiom, $AK\varphi \Leftrightarrow A\varphi$, is provable in $\mathcal{U}$, so does not have to be given separately.

structures [Heifetz, Meier, and Schipper 2005]. Although this is not the way they present it, their results actually show that $\mathcal{U}_n$ is a sound and complete axiomatization of weak validity with respect to $\mathcal{H}_n^{ret}(\Phi)$.

Thus, the following is immediate from Corollary 3.3.

Corollary 3.4: *$\mathcal{U}_n$ and S5$_n^K$ are both sound and complete axiomatization of weak validity for the language $\mathcal{L}_n^K(\Phi)$ with respect to $\mathcal{H}_n^{ret}(\Phi)$; $\mathcal{U}_n^X$ and S5$_n^X$ are both sound and complete axiomatizations of validity for the language $\mathcal{L}_n^X(\Phi)$ with respect to $\mathcal{N}_n^{ret,pd}(\Phi)$.*

We can provide a direct proof that $\mathcal{U}_n$ and S5$_n^K$ (resp., $\mathcal{U}_n^X$ and S5$_n^X$) are equivalent, without appealing to Corollary 3.3. It is easy to check that all the axioms of $\mathcal{U}_n^X$ are valid in $\mathcal{N}_n^{ret,pd}(\Phi)$ and all the inference rules of $\mathcal{U}_n^X$ preserve validity. From the completeness of S5$_n^X$ proved by Halpern, it follows that anything provable in $\mathcal{U}_n^X$ is provable in S5$_n^X$, and hence that anything provable in $\mathcal{U}_n$ is provable in S5$_n^K$. Similarly, it is easy to check that all the axioms of S5$_n^K$ are weakly valid in $\mathcal{H}_n^{ret}(\Phi)$, and the inference rules preserve validity. Thus, from the results of HMS, it follows that everything provable in S5$_n^K$ is provable in $\mathcal{U}_n$ (and hence that everything provable in S5$_n^X$ is provable in $\mathcal{U}_n^X$).[6]

These results show a tight connection between the various approaches. $\mathcal{U}$ is a sound and complete axiomatization for objective validity in partitional GSMs; $\mathcal{U}_n$ is a sound and complete axiomatization for weak validity in partitional HMS structures; and $\mathcal{U}_n^X$ is a sound and complete axiomatization for (the standard notion of) validity in partitional awareness structures where awareness is generated by primitive propositions and agents know which formulas they are aware of.

[6]The axiom systems K$_X$, T$_X$, S4$_X$, and S5$_X$ considered by Halpern [2001] include an inference rule that says that certain formulas are irrelevant in a proof. Specifically, this inference rule says that if no primitive propositions in φ appear in ψ, then from $\neg A\varphi \Rightarrow \psi$ infer ψ. Halpern conjectured that the rule was unnecessary, but could not prove it. These results show that, in the case of S5$_X$ (where we are dealing with partitional structures), the rule is in fact unnecessary, since all the axioms of $\mathcal{U}^X$ follow without it. However, its necessity for the other logics is still open.

4 Strong Validity

We say a formula is *(strongly) valid* in HMS structures if it is true at every state in every HMS structure. We can get further insight into HMS structures by considering strong validity. However, since no nontrivial formulas in $\mathcal{L}_n^K(\Phi)$ are valid in HMS structures, we must first extend the language. We do so by adding a nonstandard implication operator $\hookrightarrow$ to the language.[7] The formula $\varphi \hookrightarrow \psi$ should be read "φ implies ψ" and, as the reading suggests, we want the semantics of $\hookrightarrow$ to extend classical implication, so that if both φ and ψ are either true or false (i.e., are not undefined), then $\varphi \hookrightarrow \psi$ is true exactly if $\varphi \Rightarrow \psi$ is true in classical logic. The question is how we should define $\varphi \hookrightarrow \psi$ if one of φ or ψ is undefined. We define $\varphi \hookrightarrow \psi$ to be false under exactly the same conditions that $\varphi \Rightarrow \psi$ is taken to be false in classical logic: if φ is true and ψ is false. If φ is true, then, as we would expect, $\varphi \hookrightarrow \psi$ only if ψ is also true; if φ is undefined, then we take $\varphi \hookrightarrow \psi$ to be true, which is not that surprising. The interesting case is what happens if φ is false. In classical logic, $\varphi \Rightarrow \psi$ is automatically true in this ase. For $\hookrightarrow$, if φ is false, then $\varphi \hookrightarrow \psi$ is true only if ψ is defined. Intuitively, the reason for this is that we want $\varphi \hookrightarrow \psi$ to be valid only if (1) ψ is true whenever φ is true and (2) ψ is at least as defined as φ. Thus, although we want $(p \wedge \neg p) \hookrightarrow (p \wedge \neg p)$ to be valid, we do not want $(p \wedge \neg p) \hookrightarrow (q \wedge \neg q)$ to be valid.

Let $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ be the language that results by closing off under $\hookrightarrow$ in addition to $\neg$, $\wedge$, and $K_1, \ldots, K_n$; let $\mathcal{L}^{\hookrightarrow}(\Phi)$ be the propositional fragment of the language. We cannot use the MR definition of negation for $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$. The MR definition actually would work for formulas not involving $\hookrightarrow$, but with $\hookrightarrow$, things get more complicated. To understand why, note that if $s \in S_\Psi$ and $p \notin \Psi$, it is still the case that $p \hookrightarrow p$ is true at s (indeed, $p \hookrightarrow p$ is true at all states). We thus also want to take $\neg\neg(p \hookrightarrow p)$ to be true at s, despite the fact that $\neg(p \hookrightarrow p) \notin \mathcal{L}_n^{K,\hookrightarrow}(\Psi)$.

To deal appropriately with negation, rather than having one clause to handle the truth of $\neg\varphi$, we define

[7] We remark that a nonstandard implication operator was also added to the logic used by Fagin, Halpern, and Vardi [1995] for exactly the same reason, although the semantics of the operator here is different from there, since the underlying logic is different.

the truth of both φ and $\neg\varphi$ simultaneously by induction on the structure of φ. As usual, we say that φ is false at state s if $\neg\varphi$ is true at state s and that φ is undefined at s if it is neither true nor false at s. In the definitions, we use $(M,s) \models\uparrow\varphi$ as an abbreviation of $(M,s) \not\models \varphi$ and $(M,s) \not\models \neg\varphi$; and $(M,s) \models\downarrow\varphi$ as an abbreviation of $(M,s) \models \varphi$ or $(M,s) \models \neg\varphi$ (so $(M,s) \models\uparrow\varphi$ iff φ is neither true nor false at s, i.e., it is undefined at s). Let M be an HMS structure.

$(M,s) \models \top$

$(M,s) \not\models \neg\top$

$(M,s) \models p$ if $\pi(s,p) = 1$

$(M,s) \models \neg p$ if $\pi(s,p) = 0$

$(M,s) \models \neg\neg\varphi$ if $(M,s) \models \varphi$

$(M,s) \models \varphi \wedge \psi$ if $(M,s) \models \varphi$ and $(M,s) \models \psi$

$(M,s) \models \neg(\varphi \wedge \psi)$ if either $(M,s) \models \neg\varphi \wedge \psi$ or $(M,s) \models \varphi \wedge \neg\psi$ or $(M,s) \models \neg\varphi \wedge \neg\psi$

$(M,s) \models (\varphi \hookrightarrow \psi)$ if either $(M,s) \models \varphi \wedge \psi$ or $(M,s) \models\uparrow\varphi$ or $((M,s) \models \neg\varphi$ and $(M,s) \models\downarrow\psi)$

$(M,s) \models \neg(\varphi \hookrightarrow \psi)$ if $(M,s) \models \varphi \wedge \neg\psi$

$(M,s) \models K_i\varphi$ if $(M,s) \models\downarrow\varphi$ and $(M,t) \models \varphi$ for all $t \in \mathcal{K}_i(s)$

$(M,s) \models \neg K_i\varphi$ if $(M,s) \not\models K_i\varphi$ and $(M,s) \models\downarrow\varphi$.

It is easy to check that this semantics agrees with the MR semantics for formulas in $\mathcal{L}_n^K(\Phi)$. Moreover, the semantics for $\varphi \hookrightarrow \psi$ enforces the intuitions described earlier. In particular, it agrees with the classical semantics for $\varphi \Rightarrow \psi$ if φ and ψ are both defined; the conditions for $\varphi \hookrightarrow \psi$ to be false are exactly the same as those for $\varphi \Rightarrow \psi$ to be false, even when φ and ψ are not both defined; and if φ is false, then $\varphi \Rightarrow \psi$ is true if and only if ψ is defined. Thus, $\varphi \hookrightarrow \psi$ is valid only if ψ is at least as defined as φ.

The following lemma follows by an easy induction on the structure of formulas.

Lemma 4.1: *If $\Psi \subseteq \Psi'$, every formula in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi)$ is defined at every state in $S_{\Psi'}$.*

It is useful to define the following abbreviations:

- $\varphi \rightleftharpoons \psi$ is an abbreviation of $(\varphi \hookrightarrow \psi) \wedge (\psi \hookrightarrow \varphi)$;
- $\varphi = 1$ is an abbreviation of $\neg(\varphi \hookrightarrow \neg\top)$;
- $\varphi = 0$ is an abbreviation of $\neg(\neg\varphi \hookrightarrow \neg\top)$;
- $\varphi = \frac{1}{2}$ is an abbreviation of $(\varphi \hookrightarrow \neg\top) \wedge (\neg\varphi \hookrightarrow \neg\top)$.

Using the formulas $\varphi = 0$, $\varphi = \frac{1}{2}$, and $\varphi = 1$, we can reason directly about the truth value of formulas. This will be useful in our axiomatization.

In our axiomatization of $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ with respect to HMS structures, just as in standard epistemic logic, we focus on axioms that characterize properties of the $\mathcal{K}_i$ relation that correspond to reflexivity, transitivity, and the Euclidean property.

Consider the following axioms:

Prop$'$. All substitution instances of formulas valid in $\mathcal{L}^{\hookrightarrow}(\Phi)$.

K$'$. $K_i\varphi \wedge K_i(\varphi \hookrightarrow \psi)) \hookrightarrow K_i\psi$.

T$'$. $K_i\varphi \hookrightarrow \varphi \vee \bigvee_{\{p: p\in\Phi_\varphi\}} K_i(p = 1/2)$.

4$'$. $K_i\varphi \hookrightarrow K_iK_i\varphi$.

5$'$. $\neg K_i\neg K_i\varphi \hookrightarrow (K_i\varphi) \vee K_i(\varphi = 1/2)$.

Conf1. $(\varphi = 1/2) \hookrightarrow K_i(\varphi = 1/2)$ if $\varphi \in \mathcal{L}_n^K(\Phi)$.

Conf2. $\neg K_i(\varphi = 1/2) \hookrightarrow K_i((\varphi \vee \neg\varphi) = 1)$.

B1. $(K_i\varphi) = 1/2 \rightleftharpoons \varphi = 1/2$.

B2. $((\varphi = 0 \vee \varphi = 1) \wedge K_i(\varphi = 1)) \hookrightarrow (K_i\varphi) = 1$.

MP$'$. From φ and $\varphi \hookrightarrow \psi$ infer ψ.

A few comments regarding the axioms: Prop$'$, K$'$, T$'$, 4$'$, 5$'$, and MP$'$ are weakenings of the corresponding axioms and inference rule for standard epistemic logic. All of them use $\hookrightarrow$ rather than $\Rightarrow$; in some cases further weakening is required. We provide an axiomatic characterization of Prop$'$ in the appendix. A key property of the axiomatization is that if we just add the axiom $\varphi \neq 1/2$ (saying that all formulas are defined), we get a complete axiomatization of classical logic. T (with $\Rightarrow$ replaced by $\hookrightarrow$) is sound in HMS systems satisfying generalized reflexivity for formulas φ in $\mathcal{L}_n^K(\Phi)$. But, for example, $K_i(p = 1/2) \hookrightarrow p = 1/2$ is not valid; p may be defined (i.e., be either true or false) at a state s and undefined at all states $s' \in \mathcal{K}_i(s)$. Note that axiom 5 is equivalent to its contrapositive $\neg K_i \neg K_i\varphi \Rightarrow K_i\varphi$. This is not sound in its full strength; for example, if p is defined at s but undefined at the states in $\mathcal{K}_i(s)$, then $(M, s) \models \neg K_i \neg K_i p \wedge \neg K_i p$. Axioms Conf1 and Conf2, as the names suggest, capture confinedness. We can actually break confinedness into two parts. If $s \in S_\Psi$, the first part says that each state $s' \in \mathcal{K}_i(s)$ is in some set $S_{\Psi'}$ such that $\Psi' \subseteq \Psi$. In particular, that means that a formula in $\mathcal{L}_n^K(\Phi)$ that is undefined at s must be undefined at each state in $\mathcal{K}_i(s)$. This is just what Conf1 says. Note that Conf1 does not hold for arbitrary formulas; for example, if p is defined and q is undefined at s, and both are undefined at all states in $\mathcal{K}_i(s)$, then $(M, s) \models (p \hookrightarrow q) = 1/2 \wedge \neg K_i((p \hookrightarrow q) = 1/2)$. The second part of confinedness says that all states in $\mathcal{K}_i(s)$ are in the same set $S_{\Psi'}$. This is captured by Conf2, since it says that if φ is defined at some state in $\mathcal{K}_i(s)$, then it is defined at all states in $\mathcal{K}_i(s)$. B1 and B2 are technical axioms that capture the semantics of $K_i\varphi$.[8]

Let $\mathrm{AX}_n^{K,\hookrightarrow}$ be the system consisting of Prop$'$, K$'$, B1, B2, Conf1, Conf2, MP$'$, and Gen.

[8] We remark that axiom B2 is slightly modified from the preliminary version of the paper.

Theorem 4.2: *Let $\mathcal{C}$ be a (possibly empty) subset of $\{\mathrm{T}', 4', 5'\}$ and let C be the corresponding subset of $\{r, t, e\}$. Then $\mathrm{AX}_n^{K,\hookrightarrow} \cup \mathcal{C}$ is a sound and complete axiomatization of the language $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ with respect to $\mathcal{H}_n^C(\Phi)$.*

Theorem 4.2 also allows us to relate HMS structures to standard epistemic structures. It is easy to check that if $\mathcal{C}$ is a (possibly empty) subset of $\{\mathrm{T}', 4', 5'\}$ and C is the corresponding subset of $\{r, e, t\}$, all the axioms of $\mathrm{AX}_n^{K,\hookrightarrow} \cup \mathcal{C}$ are sound with respect to standard epistemic structures $\mathcal{M}_n^C(\Phi)$. Moreover, we get completeness by adding the axiom $\varphi \neq 1/2$, which says that all formulas are either true or false. Thus, in a precise sense, HMS differs from standard epistemic logic by allowing a third truth value.

5 Conclusion

We have compared the HMS approach and the FH approach to modeling unawareness. Our results show that, as long as we restrict attention to the language $\mathcal{L}_n^K(\Phi)$, the approaches are essentially equivalent; we can translate from one to the other. Recently, we have extended the logic of awareness by allowing awareness of unawareness [Halpern and Rêgo 2006b], so that it is possible to say, for example, that there exists a fact that agent 1 is unaware of but agent 1 knows that agent 2 is aware of it. This would be expressed by the formula $X_1 \exists x(\neg A_1 x \wedge A_2 x)$. Note that this formula is quite different from $\exists x X_1(\neg A_1 x \wedge A_2 x)$. The former formula says that agent 1 knows that there is some fact that he is not aware of but agent 2 is; however, agent 1 does not necessarily know what that fact is. According to the latter formula, there is a fact that agent 1 knows he is not aware of but agent 2 is. Thus, with the latter formula, agent 1 knows what fact he is not aware of, which cannot be the case if awareness is propositionally generated. The analogous difference arises in first-order modal logic. For example, the formula $K_1 \exists x President(x)$ says that agent 1 knows that someone is the president; $\exists x K_1 President(x)$ says that there is someone that agent 1 knows is the president. Thus, with the former formula, agent 1 does not know who the president is; he just knows that there is one. With the latter formula, he knows who the president is.

Such reasoning about awareseems of unawareness is critical to capture what is going on in a number of games. For example, taking agent 1 to be a naive investor and agent 2 to be a sophisticated broker, the investor may know that the broker is aware of some relevant facts of which he (the investor) not aware. If we take a more computational notion of awareness, then if we take agent 1 to be a novice chess player and agent 2 to be an expert, the novice may know that there are some moves that the expert is aware of which he is not (where "awareness of a move" in this setting means "understanding the consequences of the move"). See [Halpern and Rêgo 2006b] for more details of the logic. This is an instance where using the FH approach really seems to help; it is not clear whether such awawreness of unawareness can be expressed in the HMS framework.

We have also recently shown how awareness and awareness of unawareness can be added to extensive games, and have described the resulting equilibrium notions [Halpern and Rêgo 2006a]. This work is in the spirit of work on awareness in games by Feinberg [2004, 2005], but Feinberg's conditions are expressed syntactically, whereas ours are expressed semantically. We believe that there is much more to be done in understanding the impact of awareness (and lack of it) in the context of games; these are just the first steps.

Acknowledgments: We thank Aviad Heifetz, Martin Meier, and Burkhard Schipper for useful discussions on awareness. This work was supported in part by NSF under grants CTC-0208535 and ITR-0325453, by ONR under grants N00014-00-1-03-41 and N00014-01-10-511, and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the ONR under grant N00014-01-1-0795. The second author was also supported in part by a scholarship from the Brazilian Government through the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics 4*(6), 1236–1239.

Aumann, R. J. (1999). Interactive epistemology I: knowledge. *International Journal of Game Theory 28*(3), 263–301.

Bacharach, M. (1985). Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory 37*, 167–190.

Chellas, B. F. (1980). *Modal Logic*. Cambridge, U.K.: Cambridge University Press.

Dekel, E., B. Lipman, and A. Rustichini (1998). Standard state-space models preclude unawareness. *Econometrica 66*, 159–173.

Fagin, R. and J. Y. Halpern (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence 34*, 39–76.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi (1995). *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press. A revised paperback edition was published in 2003.

Fagin, R., J. Y. Halpern, and M. Y. Vardi (1995). A nonstandard approach to the logical omniscience problem. *Artificial Intelligence 74*(2), 203–240.

Feinberg, Y. (2004). Subjective reasoning—games with unawareness. Technical Report Resarch Paper Series #1875, Stanford Graduate School of Business.

Feinberg, Y. (2005). Games with incomplete awareness. Technical Report Resarch Paper Series #1894, Stanford Graduate School of Business.

Geanakoplos, J. (1989). Game theory without partitions, and applications to speculation and consensus. Cowles Foundation Discussion Paper #914, Yale University.

Halpern, J. Y. (2001). Alternative semantics for unawareness. *Games and Economic Behavior 37*, 321–339.

Halpern, J. Y. and L. C. Rêgo (2006a). Extensive games with possibly unaware players. In *Proc. Fifth*

International Joint Conference on Autonomous Agents and Multiagent Systems.

Halpern, J. Y. and L. C. Rêgo (2006b). Reasoning about knowledge of unawareness. In *Principles of Knowledge Representation and Reasoning: Proc. Tenth International Conference (KR '06)*. Full version available at arxiv.org/cs.LO/0603020.

Heifetz, A., M. Meier, and B. Schipper (2003). Multi-person unawareness. In *Theoretical Aspects of Rationality and Knowledge: Proc. Ninth Conference (TARK 2003)*, pp. 148–158. An extended version with the title "Interactive unawareness" will appear in the *Journal of Economic Theory*.

Heifetz, A., M. Meier, and B. Schipper (2005). A canonical model for interactive unawareness. Unpublished manuscript.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

Modica, S. and A. Rustichini (1994). Awareness and partitional information structures. *Theory and Decision 37*, 107–124.

Modica, S. and A. Rustichini (1999). Unawareness and partitional information structures. *Games and Economic Behavior 27*(2), 265–298.

Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. Cambridge, Mass.: MIT Press.

Samet, D. (1987). Ignoring ignorance and agreeing to disagree. MEDS discussion paper, Northwestern University.

A The Original HMS Approach

HMS describe their approach purely semantically, without giving a logic. We review their approach here (making some inessential changes for ease of exposition). An *HMS frame for n agents* is a tuple $(\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, (\Delta, \preceq), \{\rho_{\beta,\alpha} : \alpha, \beta \in \Delta, \alpha \preceq \beta\})$, where:

- Δ is an arbitrary lattice, partially ordered by $\preceq$;

- $\mathcal{K}_1, \ldots, \mathcal{K}_n$ are possibility correspondences, one for each agent;

- Σ is a disjoint union of the form $\cup_{\alpha \in \Delta} S_\alpha$;

- if $\alpha \preceq \beta$, then $\rho_{\beta,\alpha} : S_\beta \rightarrow S_\alpha$ is a surjection.

In the logic-based version of HMS given in Section 2.4, Δ consists of the subsets of Φ, and $\Psi \preceq \Psi'$ iff $\Psi \subseteq \Psi'$. Thus, the original HMS definition can be viewed as a more abstract version of that given in Section 2.4.

Given $B \subseteq S_\alpha$, let $B^\uparrow = \cup_{\{\beta:\, \alpha \preceq \beta\}} \rho_{\beta,\alpha}^{-1}(B)$. Again, $B^\uparrow$ consists of the states that project to a state in B. HMS focus on sets of the form $B^\uparrow$, which they take to be events.

HMS assume that their frames satisfy the five conditions mentioned in Section 2.4, restated in their more abstract setting. The statements of generalized reflexivity and stationarity remain the same. Confinedness, projections preserve knowledge, and projections preserve ignorance are stated as follows:

- confinedness: if $s \in S_\beta$ then $\mathcal{K}_i(s) \subseteq S_\alpha$ for some $\alpha \preceq \beta$;

- projections preserve knowledge: if $\alpha \preceq \beta \preceq \gamma$, $s \in S_\gamma$, and $\mathcal{K}_i(s) \subseteq S_\beta$, then $\rho_{\beta,\alpha}(\mathcal{K}_i(s)) = \mathcal{K}_i(\rho_{\gamma,\alpha}(s))$;

- projections preserve ignorance: if $s \in S_\beta$ and $\alpha \preceq \beta$ then $(\mathcal{K}_i(s))^\uparrow \subseteq (\mathcal{K}_i(\rho_{\beta,\alpha}(s)))^\uparrow$.

HMS start by considering the algebra consisting of all events of the form $B^\uparrow$. In this algebra, they define an operator $\neg$ by taking $\neg(B^\uparrow) = (S_\alpha - B)^\uparrow$ for $\emptyset \neq B \subseteq S_\alpha$. With this definition, $\neg\neg B^\uparrow = B^\uparrow$ if $B \notin \{\emptyset, S_\alpha\}$. However, it remains to define $\neg\emptyset^\uparrow$. We could just take it to be Σ, but then we have $\neg\neg S_\alpha^\uparrow = \Sigma$, rather than $\neg\neg S_\alpha^\uparrow = S_\alpha^\uparrow$. To avoid this problem, in their words, HMS "devise a distinct vacuous event $\emptyset^{S_\alpha}$" for each subspace S_α, extend the algebra with these events, and define $\neg S_\alpha^\uparrow = \emptyset^{S_\alpha}$ and $\neg\emptyset^{S_\alpha} = S_\alpha^\uparrow$. They do not make clear exactly what it means to "devise a vacuous event". We can recast their definitions in the following way, that allows us to bring in the events $\emptyset^{S_\alpha}$ more naturally.

In a 2-valued logic, given a formula φ and a structure M, the set $[\![\varphi]\!]_M$ of states where φ is true and the set $[\![\neg\varphi]\!]_M$ of states where φ is false are complements of each other, so it suffices to associate with φ only one set, say $[\![\varphi]\!]_M$. In a 3-valued logic, the set of states where φ is true does not determine the set of states where φ is false. Rather, we must consider three mutually exclusive and exhaustive sets: the set where φ is true, the set where φ is false, and the set where φ is undefined. As before, one of these is redundant, since it is the complement of the union of the other two. Note that if φ is a formula in the language of HMS, the set $[\![\varphi]\!]_M$ is either $\emptyset$ or an event of the form $B^{\uparrow}$, where $B \subseteq S_\alpha$. In the latter case, we associate with φ the pair of sets $(B^{\uparrow}, (S_\alpha - B)^{\uparrow})$, i.e., $([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M)$. In the former case, we must have $[\![\neg\varphi]\!]_M = S_\alpha^{\uparrow}$ for some α, and we associate with φ the pair $(\emptyset, S_\alpha^{\uparrow})$. Thus, we are using the pair $(\emptyset, S_\alpha^{\uparrow})$ instead of devising a new event $\emptyset^{S_\alpha}$ to represent $[\![\varphi]\!]_M$ in this case.[9]

HMS use intersection of events to represent conjunction. It is not hard to see that the intersection of events is itself an event. The obvious way to represent disjunction is as the union of events, but the union of events is in general not an event. Thus, HMS define a disjunction operator using de Morgan's law: $E \vee E' = \neg(\neg E \cap \neg E')$. In our setting, where we use pairs of sets, we can also define operators $\sim$ and $\sqcap$ (intuitively, for negation and intersection) by taking $\sim(E, E') = (E', E)$ and

$$(E, E') \sqcap (F, F') = (E \cap F, (E \cap F') \cup (E' \cap F) \cup (E' \cap F')).$$

Although our definition of $\sqcap$ may not seem so intuitive, as the next result shows, $(E, E') \sqcap (F, F')$ is essentially equal to $(E \cap F, \neg(E \cap F))$. Moreover, our definition has the advantage of not using $\neg$, so it applies even if E and F are not events.

[9] In a more recent version of their paper, HMS identify a nonempty event E with the pair (E, S), where, for $E = B^{\uparrow}$, S is the unique set S_α containing B. Then $\emptyset^S$ can be identified with $(\emptyset, S)$. While we also identify events with pairs of sets and $\emptyset^S$ with $(\emptyset, S)$, our identification is different from that of HMS, and extends more naturally to sets that are not events.

Lemma A.1: *If* $(E \cup E') = S_\alpha^\uparrow$ *and* $(F \cup F') = S_\beta^\uparrow$, *then*

$$(E \cap F') \cup (E' \cap F) \cup (E' \cap F') = \begin{cases} \neg(E \cap F) & \text{if } (E \cap F) \neq \emptyset, \\ S_\gamma^\uparrow & \text{if } (E \cap F) = \emptyset \text{ and } \gamma = \sup(\alpha, \beta).^{10} \end{cases}$$

Proof: Let I be the set $(E \cap F) \cup (E \cap F') \cup (E' \cap F) \cup (E' \cap F')$. We first show that $I = S_\gamma^\uparrow$, where $\gamma = \sup(\alpha, \beta)$. By assumption, $E = B^\uparrow$ for some $B \subseteq S_\alpha$, and $F = C^\uparrow$ for some $C \subseteq S_\beta$. Suppose that $s \in I$. We claim that $s \in S_\delta$, where $\alpha \preceq \delta$ and $\beta \preceq \delta$. Suppose, by way of contradiction, that $\alpha \not\preceq \delta$. Then $s \notin E \cup E'$, so $s \notin I$, a contradiction. A similar argument shows that $\beta \preceq \delta$. Thus $\gamma \preceq \delta$ and $s \in S_\gamma^\uparrow$. For the opposite inclusion, suppose that $s \in S_\gamma^\uparrow$. Since $\alpha \preceq \gamma$ and $\beta \preceq \gamma$, the projections $\rho_{\gamma,\alpha}(s)$ and $\rho_{\gamma,\beta}(s)$ are well defined. Let $X = E$ if $\rho_{\gamma,\alpha}(s) \in B$ and $X = E'$ otherwise. Similarly, let $Y = F$ if $\rho_{\gamma,\beta}(s) \in C$ and $Y = F'$ otherwise. It is easy to see that $s \in (X \cap Y) \subseteq I$. It follows that $(E \cap F') \cup (E' \cap F) \cup (E' \cap F') = S_\gamma^\uparrow - (E \cap F)$. The result now follows easily. ∎

Finally, HMS define an operator K_i corresponding to the possibility correspondence $\mathcal{K}_i$. They define $\mathsf{K}_i(E) = \{s : \mathcal{K}_i(s) \subseteq E\}$,[11] and show that $\mathsf{K}_i(E)$ is an event if E is. In our setting, we define

$$\mathsf{K}_i((E, E')) = (\{s : \mathcal{K}_i(s) \subseteq E\} \cap (E \cup E'), (E \cup E') - \{s : \mathcal{K}_i(s) \subseteq E\}).$$

Essentially, we are defining $\mathsf{K}_i((E, E'))$ as $(\mathsf{K}_i(E), \neg\mathsf{K}_i(E))$. Intersecting with $E \cup E'$ is unnecessary in the HMS framework, since their conditions on frames guarantee that $\mathsf{K}_i(E) \subseteq E \cup E'$. If we think of (E, E') as $([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M)$, then φ is defined on $E \cup E'$. The definitions above guarantee that $K_i\varphi$ is defined on the same set.

[10] Note that $\sup(\alpha, \beta)$ is well defined since Δ is a lattice.

[11] Actually, this is their definition only if $\{s : \mathcal{K}_i(s) \subseteq E\} \neq \emptyset$; otherwise, they take $\mathsf{K}_i(E) = \emptyset^{S_\alpha}$ if $E = B^\uparrow$ for some $B \subseteq S_\alpha$. We do not need a special definition if $\{s : \mathcal{K}_i(s) \subseteq E\} = \emptyset$ using our approach.

HMS define an awareness operator in the spirit of MR, by taking $\mathsf{A}_i(E)$ to be an abbreviation of $\mathsf{K}_i(E) \vee \mathsf{K}_i \neg \mathsf{K}_i(E)$. They then prove a number of properties of knowledge and awareness, such as $\mathsf{K}_i(E) \subseteq \mathsf{K}_i\mathsf{K}_i(E)$ and $\mathsf{A}_i(\neg E) = \mathsf{A}_i(E)$.

The semantics we have given for our logic matches that of the operators defined by HMS, in the sense of the following lemma.

Lemma A.2: *For all formulas $\varphi, \psi \in \mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ and all HMS structures M.*

(a) $([\![\neg\varphi]\!]_M, [\![\neg\neg\varphi]\!]_M) = \sim([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M))$

(b) $([\![\varphi \wedge \psi]\!]_M, [\![\neg(\varphi \wedge \psi)]\!]_M) = ([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M) \sqcap ([\![\psi]\!]_M, [\![\neg\psi]\!]_M)$.

(c) $([\![K_i\varphi]\!]_M, [\![\neg K_i\varphi]\!]_M) = \mathsf{K}_i(([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M))$

Proof: Part (a) follows easily from the fact that $\sim([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M)) = ([\![\neg\varphi]\!]_M), [\![\varphi]\!]_M)$ and $[\![\neg\neg\varphi]\!]_M = [\![\varphi]\!]_M$.

For part (b), note that

$$([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M) \sqcap ([\![\psi]\!]_M, [\![\neg\psi]\!]_M) = ([\![\varphi]\!]_M \cap [\![\psi]\!]_M, ([\![\varphi]\!]_M \cap [\![\neg\psi]\!]_M) \cup ([\![\neg\varphi]\!]_M \cap [\![\psi]\!]_M) \cup ([\![\neg\varphi]\!]_M \cap [\![\neg\psi]\!]_M)).$$

Now the result is immediate from the observation that $[\![\varphi]\!]_M \cap [\![\psi]\!]_M = [\![\varphi \wedge \psi]\!]_M$ and

$$([\![\varphi]\!]_M \cap [\![\neg\psi]\!]_M) \cup ([\![\neg\varphi]\!]_M \cap [\![\psi]\!]_M) \cup ([\![\neg\varphi]\!]_M \cap [\![\neg\psi]\!]_M) = [\![\neg(\varphi \wedge \psi)]\!]_M.$$

For (c), by definition of K_i,

$$\mathsf{K}_i(([\![\varphi]\!]_M, [\![\neg\varphi]\!]_M)) = (\{s : \mathcal{K}_i(s) \subseteq [\![\varphi]\!]_M\} \cap ([\![\varphi]\!]_M \cup [\![\neg\varphi]\!]_M), ([\![\varphi]\!]_M \cup [\![\neg\varphi]\!]_M) - \{s : \mathcal{K}_i(s) \subseteq [\![\varphi]\!]_M\}).$$

Note that $t \in ([\![\varphi]\!]_M \cup [\![\neg\varphi]\!]_M)$ iff $(M, t) \models\downarrow \varphi$, and $t \in \{s : \mathcal{K}_i(s) \subseteq [\![\varphi]\!]_M\}$ iff for all $t' \in \mathcal{K}_i(t)$, $(M, t') \models \varphi$. Thus, $t \in \{s : \mathcal{K}_i(s) \subseteq [\![\varphi]\!]_M\} \cap ([\![\varphi]\!]_M \cup [\![\neg\varphi]\!]_M)$ iff $(M, t) \models K_i\varphi$, i.e., iff $t \in [\![K_i\varphi]\!]_M$. Similarly, $t \in ([\![\varphi]\!]_M \cup [\![\neg\varphi]\!]_M) - \{s : \mathcal{K}_i(s) \subseteq [\![\varphi]\!]_M\}$ iff $(M, t) \models\downarrow \varphi$ and $(M, t) \not\models K_i\varphi$, i.e., iff $(M, t) \models \neg K_i\varphi$. Hence, $t \in [\![\neg K_i\varphi]\!]_M$. ∎

Note that Lemma A.2 applies even though, once we introduce the $\hookrightarrow$ operator, $[\![\varphi]\!]_M$ is not in general an event in the HMS sense. (For example, $[\![p \hookrightarrow q]\!]_M$ is not in general an event.)

B An Axiomatization of $\mathcal{L}^{\hookrightarrow}(\Phi)$

Note that the formulas $\varphi = 0$, $\varphi = \frac{1}{2}$, and $\varphi = 1$ are 2-valued. More generally, we define a formula φ to be *2-valued* if $(\varphi = 0) \vee (\varphi = 1)$ is valid in all HMS structures. Because they obey the usual axioms of classical logic, 2-valued formulas play a key role in our axiomatization of $\mathcal{L}^{\hookrightarrow}(\Phi)$. We say that a formula is *definitely 2-valued* if it is in the smallest set containing $\top$ and all formulas of the form $\varphi = k$ which is closed under negation, conjunction, nonstandard implication, and K_i, so that if φ and ψ are definitely two-valued, then so are $\neg\varphi$, $\varphi \wedge \psi$, $\varphi' \hookrightarrow \psi$ (for all φ'), and $K_i\varphi$. Let D_2 denote the set of definitely 2-valued formulas.

The following lemma is easy to prove.

Lemma B.1: *If φ is definitely 2-valued, then it is 2-valued.*

Let AX_3 consist of the following collection of axioms and inference rules:

P0. $\top$.

P1. $(\varphi \wedge \psi) \rightleftharpoons \neg(\varphi \hookrightarrow \neg\psi)$ if $\varphi, \psi \in D_2$.

P2. $\varphi \hookrightarrow (\psi \hookrightarrow \varphi)$ if $\varphi, \psi \in D_2$.

P3. $(\varphi \hookrightarrow (\psi \hookrightarrow \varphi')) \hookrightarrow ((\varphi \hookrightarrow \psi) \hookrightarrow (\varphi \hookrightarrow \varphi'))$ if $\varphi, \psi, \varphi' \in D_2$.

P4. $(\varphi \hookrightarrow \psi) \hookrightarrow ((\varphi \hookrightarrow \neg\psi) \hookrightarrow \neg\varphi)$ if $\varphi, \psi \in D_2$.

P5. $(\varphi \wedge \psi) = 1 \rightleftharpoons (\varphi = 1) \wedge (\psi = 1)$.

P6. $(\varphi \wedge \psi) = 0 \rightleftharpoons (\varphi = 0 \wedge \neg(\psi = 1/2)) \vee (\neg(\varphi = 1/2) \wedge \psi = 0)$.

P7. $\varphi = 1 \rightleftharpoons (\neg\varphi) = 0$.

P8. $\varphi = 0 \rightleftharpoons (\neg\varphi) = 1$.

P9. $(\varphi \hookrightarrow \psi) = 1 \rightleftharpoons ((\varphi = 0 \wedge \neg(\psi = 1/2)) \vee (\varphi = 1/2) \vee (\varphi = 1 \wedge \psi = 1))$.

P10. $(\varphi \hookrightarrow \psi) = 0 \rightleftharpoons (\varphi = 1 \wedge \psi = 0)$.

P11. $(\varphi = 0 \vee \varphi = 1/2 \vee \varphi = 1) \wedge (\neg(\varphi = i \wedge \varphi = j))$, for $i, j \in \{0, 1/2, 1\}$ and $i \neq j$.

R1. From $\varphi = 1$ infer φ.

MP$'$. From φ and $\varphi \hookrightarrow \psi$ infer ψ.

It is well known that P0-P4 together with MP$'$ provide a complete axiomatization for classical 2-valued propositional logic with negation, conjunction, implication, and $\top$.[12] Axioms P5-P10 are basically a translation to formulas of the semantics for conjunction, negation and implication.

Note that all the axioms of AX_3 are sound in classical logic (all formulas of the form $\varphi = 1/2$ are vacuously false in classical logic). Moreover, it is easy to show that if we add the axiom $\neg(\varphi = 1/2)$ to AX_3, we get a sound and complete axiomatization of classical propositional logic (although many axioms then become redundant).

Theorem B.2: *AX_3 is a sound and complete axiomatization of $\mathcal{L}^{\hookrightarrow}(\Phi)$.*

Proof: The proof that the axiomatization is sound is a straightforward induction on the length of the proof of any theorem φ. We omit the details here. For completeness, we need to show that a valid formula $\varphi \in \mathcal{L}^{\hookrightarrow}(\Phi)$ is provable in AX_3. We first prove that $\varphi = 1$ is provable in AX_3 using standard techniques, and then apply R1 to infer φ. We proceed as follows.

Given a set $\mathcal{G}$ of formulas, let $\wedge\mathcal{G} = \bigwedge_{\varphi \in \mathcal{G}} \varphi$. A set $\mathcal{G}$ of formulas is *AX-consistent*, if for all finite subsets $\mathcal{G}' \subseteq \mathcal{G}$, AX $\not\vdash \neg(\wedge\mathcal{G}')$. A set $\mathcal{G}$ of formulas is *maximal AX-consistent* if $\mathcal{G}$ is AX-consistent and for all $\varphi \notin \mathcal{G}$, $\mathcal{G} \cup \{\varphi\}$ is not AX-consistent.

[12] We remark that we included formulas of the form $K_i\varphi$ among the formulas that are definitely 2-valued. While such formulas are not relevant in the axiomatization of $\mathcal{L}^{\hookrightarrow}(\Phi)$, they do play a role when we consider the axiom Prop$'$ in $\mathrm{AX}_n^{K,\hookrightarrow}$, which applies to instances in the language $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ of valid formulas of $\mathcal{L}^{\hookrightarrow}(\Phi)$.

Lemma B.3: *If $\mathcal{G}$ is an AX-consistent subset of $\mathcal{G}'$, then $\mathcal{G}$ can be extended to a maximal AX-consistent subset of $\mathcal{G}'$.*

Proof: The proof uses standard techniques. Let $\psi_1, \psi_2, \ldots$ be an enumeration of the formulas in $\mathcal{G}'$. Define $F_0 = \mathcal{G}$ and $F_i = F_{i-1} \cup \{\psi_i\}$ if $F_{i-1} \cup \{\psi_i\}$ is AX-consistent and $F_i = F_{i-1}$, otherwise. Let $\mathcal{F} = \cup_{i=0}^{\infty} F_i$. We claim that $\mathcal{F}$ is an maximal AX-consistent subset of $\mathcal{G}'$. Suppose that $\psi \in \mathcal{G}'$ and $\psi \notin \mathcal{F}$. By construction, we have $\psi = \psi_k$ for some k. If $F_{k-1} \cup \{\psi_k\}$ were AX-consistent, then ψ_k would be in F_k and hence ψ_k would be in $\mathcal{F}$. Since $\psi_k = \psi \notin \mathcal{F}$, we have that $F_{k-1} \cup \{\psi\}$ is not AX-consistent and hence $\mathcal{F} \cup \{\psi\}$, is not AX-consistent. ∎

The next lemma shows that maximal AX_3-consistent sets of definitely 2-valued formulas satisfy essentially the same properties as maximal classically consistent sets of formulas.

Lemma B.4: *Let AX be any axiom system that includes AX_3. For all maximal AX-consistent subsets $\mathcal{F}$ of D_2, the following properties hold:*

(1) for every formula $\varphi \in D_2$, exactly one of φ and $\neg\varphi$ is in $\mathcal{F}$;

(2) for every formula $\varphi \in \mathcal{L}^{\hookrightarrow}(\Phi)$, exactly one of $\varphi = 0$, $\varphi = 1/2$, and $\varphi = 1$ is in $\mathcal{F}$;

(3) if $\varphi_1, \ldots, \varphi_k, \psi \in D_2$, $\varphi_1, \ldots, \varphi_k \in \mathcal{F}$, and $\mathrm{AX}_3 \vdash (\varphi_1 \wedge \ldots \wedge \varphi_k) \hookrightarrow \psi$, then $\psi \in \mathcal{F}$;

(4) $(\varphi \wedge \psi) = 1 \in \mathcal{F}$ iff $\varphi = 1 \in \mathcal{F}$ and $\psi = 1 \in \mathcal{F}$;

(5) $(\varphi \wedge \psi) = 0 \in \mathcal{F}$ iff either $\varphi = 0 \in \mathcal{F}$ and $\psi = 1/2 \notin \mathcal{F}$, or $\psi = 0 \in \mathcal{F}$ and $\varphi = 1/2 \notin \mathcal{F}$;

(6) $\psi = 1 \in \mathcal{F}$ iff $(\neg\psi) = 0 \in \mathcal{F}$;

(7) $\psi = 0 \in \mathcal{F}$ iff $(\neg\psi) = 1 \in \mathcal{F}$;

(8) $(\varphi \hookrightarrow \psi) = 1 \in \mathcal{F}$ iff either $\varphi = 0 \in \mathcal{F}$ and $\psi = 1/2 \notin \mathcal{F}$; or $\varphi = 1/2 \in \mathcal{F}$; or $\varphi = 1 \in \mathcal{F}$ and $\psi = 1 \in \mathcal{F}$;

(9) $(\varphi \hookrightarrow \psi) = 0 \in \mathcal{F}$ *iff* $\varphi = 1 \in \mathcal{F}$ *and* $\psi = 0 \in \mathcal{F}$*;*

(10) if $\varphi \in D_2$ *and* $\mathrm{AX} \vdash \varphi$*, then* $\varphi \in \mathcal{F}$*;*

Proof: First, note that axioms P0-P4 and MP$'$ guarantee that classical propositional reasoning can be used for formulas in D_2. We thus use classical propositional reasoning with minimal comment.

For (1), we first show that exactly one of $\mathcal{F} \cup \{\varphi\}$ and $\mathcal{F} \cup \{\neg\varphi\}$ is AX-consistent. Suppose that $\mathcal{F} \cup \{\varphi\}$ and $\mathcal{F} \cup \{\neg\varphi\}$ are both AX-consistent. Then $\varphi \in \mathcal{F}$ and $\neg\varphi \in \mathcal{F}$. Since $\neg(\varphi \wedge \neg\varphi) \in D_2$, $\mathrm{AX} \vdash \neg(\varphi \wedge \neg\varphi)$, but this is a contradiction since $\mathcal{F}$ is AX-consistent. Now suppose that neither $\mathcal{F} \cup \{\varphi\}$ nor $\mathcal{F} \cup \{\neg\varphi\}$ is AX-consistent. Then there exist finite subsets $H_1, H_2 \subseteq \mathcal{F}$ such that $\mathrm{AX} \vdash \neg(\varphi \wedge (\wedge H_1))$ and $\mathrm{AX} \vdash \neg(\neg\varphi \wedge (\wedge H_2))$. Let $G = H_1 \cup H_2$. By classical propositional reasoning, $\mathrm{AX} \vdash \neg(\varphi \wedge (\wedge G))$ and $\mathrm{AX} \vdash \neg(\neg\varphi \wedge (\wedge G))$, so $\mathrm{AX} \vdash \neg((\varphi \wedge (\wedge G)) \vee (\neg\varphi \wedge (\wedge G)))$ and $\mathrm{AX} \vdash \neg((\varphi \wedge (\wedge G)) \vee (\neg\varphi \wedge (\wedge G))) \hookrightarrow \neg(\wedge G)$. Hence, by MP$'$, $\mathrm{AX} \vdash \neg(\wedge G)$. This is a contradiction, since $G \subseteq \mathcal{F}$ and $\mathcal{F}$ is AX-consistent.

Suppose that $\mathcal{F} \cup \{\varphi\}$ is AX-consistent (the other case is completely analogous). Since $\mathcal{F}$ is a maximal AX-consistent subset of D_2 and $\varphi \in D_2$, we have $\varphi \in \mathcal{F}$. And since $\mathcal{F} \cup \{\neg\varphi\}$ is not AX-consistent, $\neg\varphi \notin \mathcal{F}$.

For (2), we first show that exactly one of $\mathcal{F} \cup \{\varphi = i\}$, for $i = \{0, 1/2, 1\}$, is AX-consistent. Suppose that $\mathcal{F} \cup \{\varphi = i\}$ and $\mathcal{F} \cup \{\varphi = j\}$, $i \neq j$, are AX-consistent. Then $\varphi = i \in \mathcal{F}$ and $\varphi = j \in \mathcal{F}$. By axiom P11, $\mathrm{AX} \vdash \neg(\varphi = i \wedge \varphi = j)$. This is a contradiction, since $\mathcal{F}$ is AX-consistent.

Next, suppose that none of $\mathcal{F} \cup \{\varphi = i\}$ is AX-consistent. Then there exist finite sets $F_i \subseteq \mathcal{F}$ such that $\mathrm{AX} \vdash \neg(\varphi = i \wedge (\wedge F_i))$, $i = 0, 1/2, 1$. Let $G = F_0 \cup F_{1/2} \cup F_1$. By classical propositional reasoning, $\mathrm{AX} \vdash \neg(\varphi = i \wedge (\wedge G))$, and $\mathrm{AX} \vdash \neg((\varphi = 0 \wedge (\wedge G)) \vee (\varphi = 1/2 \wedge (\wedge G)) \vee (\varphi = 1 \wedge (\wedge G)))$. Now using axiom P11, we have $\mathrm{AX} \vdash \neg(\wedge G)$. This is a contradiction, since $G \subseteq \mathcal{F}$ and $\mathcal{F}$ is AX-consistent.

Let i^* be the unique i such that $\mathcal{F} \cup \{\varphi = i^*\}$ is AX-consistent. Since $\mathcal{F}$ is a maximal AX-consistent subset of D_2 and $\varphi = i^* \in D_2$, we have that $\{\varphi = i^*\} \in \mathcal{F}$. And since $\mathcal{F}$ is AX-consistent, it is clear by

P11, that if $j \neq i^*$, then $\{\varphi = j\} \notin \mathcal{F}$.

For (3), by part (1), if $\psi \notin \mathcal{F}$, then $\neg\psi \in \mathcal{F}$. Thus, $\{\varphi_1, \ldots, \varphi_k, \neg\psi\} \subseteq \mathcal{F}$. But since $\mathrm{AX}_3 \vdash \varphi_1 \wedge \ldots \wedge \varphi_k \hookrightarrow \psi$, by classical propositional reasoning, $\mathrm{AX}_3 \vdash \neg(\varphi_1 \wedge \ldots \wedge \varphi_k \wedge \neg\psi)$, a contradiction since $\mathcal{F}$ is AX-consistent.

The proof of the remaining properties follows easily from parts (2) and (3). For example, for part (4), if $(\varphi \wedge \psi) = 1 \in \mathcal{F}$, then the fact that $\varphi = 1 \in \mathcal{F}$ and $\psi = 1 \in F$ follows from P5 and (3). We leave details to the reader. ∎

A formula φ is said to be *satisfiable in a structure* M if $(M, s) \models \varphi$ for some world in M; φ is *satisfiable in a class of structures* $\mathcal{N}$ if it is satisfiable in at least one structure in $\mathcal{N}$. Let $\mathcal{M}_P$ be the class of all 3-valued propositional HMS models.

Lemma B.5: *If $\varphi = i$ is AX$_3$-consistent, then $\varphi = i$ is satisfiable in $\mathcal{M}_P$, for $i \in \{0, 1/2, 1\}$.*

Proof: We construct a special model $M^c \in \mathcal{M}_P$ called the *canonical 3-valued model*. M^c has a state s_V corresponding to every V that is a maximal AX$_3$-consistent subset of D_2. We show that

$$(M^c, s_V) \models \varphi = j \text{ iff } \varphi = j \in V, \text{ for } j \in \{0, 1/2, 1\}.$$

Note that this claim suffices to prove Lemma B.5 since, by Lemma B.3, if $\varphi = i$ is AX$_3$-consistent, then it is contained in a maximal AX$_3$-consistent subset of D_2. We proceed as follows. Let $M^c = (\Sigma, \pi)$, where $\Sigma = \{s_V : V \text{ is a maximal consistent subset of } D_2\}$ and

$$\pi(s_V, p) = \begin{cases} 1 & \text{if } p = 1 \in V, \\ 0 & \text{if } p = 0 \in V, \\ 1/2 & \text{if } p = 1/2 \in V. \end{cases}$$

Note that by Lemma B.4(2), the interpretation π is well defined.

We now show that the claim holds by induction on the structure of formulas. If ψ is a primitive proposition, this follows from the definition of $\pi(s_V, \psi)$.

Suppose that $\psi = \neg\varphi$. By Lemma B.4(7), $(\neg\varphi) = 1 \in V$ iff $\varphi = 0 \in V$. By the induction hypothesis, $\varphi = 0 \in V$ iff $(M^c, s_V) \models \varphi = 0$. By the semantics of the logic, we have $(M^c, s_V) \models \varphi = 0$ iff $(M^c, s_V) \models \neg\varphi$, and the latter holds iff $(M^c, s_V) \models (\neg\varphi) = 1$. Similarly, using Lemma B.4(6), we can show $(\neg\varphi) = 0 \in V$ iff $(M^c, s_V) \models (\neg\varphi) = 0$. The remaining case $\varphi = 1/2$ follows from the previous cases, axiom P11, and the fact that $(M^c, s_V) \models \psi = i$ for exactly one $i \in \{0, 1/2, 1\}$. (For all the following steps of the induction the case $\varphi = 1/2$ is omitted since it follows from the other cases for exactly the same reason.)

Suppose that $\psi = \varphi_1 \wedge \varphi_2$. By Lemma B.4(4), $\psi = 1 \in V$ iff $\varphi_1 = 1 \in V$ and $\varphi_2 = 1 \in V$. By the induction hypothesis, $\varphi_j = 1 \in V$ iff $(M^c, s_V) \models \varphi_j = 1$ for $j \in 1, 2$, which is true iff $(M^c, s_V) \models (\varphi_1 \wedge \varphi_2) = 1$. Similarly, using Lemma B.4(5), we can show that $(\varphi_1 \wedge \varphi_2) = 0 \in V$ iff $(M^c, s_V) \models (\varphi_1 \wedge \varphi_2) = 0$.

Suppose that $\psi = \varphi_1 \hookrightarrow \varphi_2$. By Lemma B.4(8), $\psi = 1 \in V$ iff either $\varphi_1 = 0 \in V$ and $\varphi_2 = 1/2 \notin V$; or $\varphi_1 = 1/2 \in V$; or $\varphi_1 = 1 \in V$ and $\varphi_2 = 1 \in V$. By the induction hypothesis, this is true iff either $(M^c, s_V) \models \varphi_1 = 0$ and $(M^c, s_V) \not\models \varphi_2 = 1/2$; or $(M^c, s_V) \models \varphi_1 = 1/2$; or $(M^c, s_V) \models \varphi_1 = 1$ and $(M^c, s_V) \models \varphi_2 = 1$. This, in turn, is true iff $(M^c, s_V) \models \neg\varphi_1$ and $(M^c, s_V) \models \neg(\varphi_2 = 1/2)$; or $(M^c, s_V) \models (\varphi_1 = 1/2)$; or $(M^c, s_V) \models \varphi_1$ and $(M^c, s_V) \models \varphi_2$. By the semantics of $\hookrightarrow$, this holds iff $(M^c, s_V) \models (\varphi \hookrightarrow \psi) = 1$. Similarly, using Lemma B.4(9), we can show that $(\varphi \hookrightarrow \psi) = 0 \in V$ iff $(M^c, s_V) \models (\varphi \hookrightarrow \psi) = 0$. ∎

We can finally complete the proof of Theorem B.2. Suppose that φ is valid. This implies $(\varphi = 0) \vee (\varphi = 1/2)$ is not satisfiable. By Lemma B.5, $(\varphi = 0) \vee (\varphi = 1/2)$ is not AX$_3$-consistent, so AX$_3 \vdash \neg((\varphi = 0) \vee (\varphi = 1/2))$. By axioms P0-P4, P11 and MP$'$, AX$_3 \vdash \varphi = 1$. And finally, applying R1, AX$_3 \vdash \varphi$. So, the axiomatization is complete. ∎

C Proofs of Theorems

In this section, we provide proofs of the theorems in Sections 3 and 4. We restate the results for the reader's convenience.

The next lemma, which is easily proved by induction on the structure of formulas, will be used throughout. We leave the proof to the reader.

Lemma C.1: *If $M \in H_n(\Phi)$, $\Psi' \subseteq \Psi \subseteq \Phi$, $s \in \Psi$, $s' = \rho_{\Psi,\Psi'}(s)$, $\varphi \in \mathcal{L}_n^K(\Phi)$, and $(M, s') \models \varphi$, then $(M, s) \models \varphi$.*

Theorem 3.2: *Let C be a subset of $\{r, t, e\}$.*

(a) If $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\}) \in \mathcal{H}_n^C(\Phi)$, then there exists an awareness structure $M' = (\Sigma, \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi', \mathcal{A}_1, \ldots, \mathcal{A}_n) \in \mathcal{N}_n^{C,pg}(\Phi)$ such that, for all $\varphi \in \mathcal{L}_n^K(\Phi)$, if $s \in S_\Psi$ and $\Phi_\varphi \subseteq \Psi$, then $(M, s) \models \varphi$ iff $(M', s) \models \varphi_X$. Moreover, if $C \cap \{t, e\} \neq \emptyset$, then we can take $M' \in \mathcal{N}_n^{C,pd}$.

(b) If $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \mathcal{A}_1, \ldots, \mathcal{A}_n) \in \mathcal{N}_n^{C,pd}(\Phi)$, then there exists an HMS structure $M' = (\Sigma', \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi', \{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\}) \in \mathcal{H}_n^C(\Phi)$ such that $\Sigma' = \Sigma \times 2^\Phi$, $S_\Psi = \Sigma \times \{\Psi\}$ for all $\Psi \subseteq \Phi$, and, for all $\varphi \in \mathcal{L}_n^K(\Phi)$, if $\Phi_\varphi \subseteq \Psi$, then $(M, s) \models \varphi_X$ iff $(M', (s, \Psi)) \models \varphi$. If $\{t, e\} \cap C = \emptyset$, then the result holds even if $M \in (\mathcal{N}_n^{C,pg}(\Phi) - \mathcal{N}_n^{C,pd}(\Phi))$.

Proof: For part (a), given $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\}) \in \mathcal{H}_n^C(\Phi)$, let $M' = (\Sigma, \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi', \mathcal{A}_1, \ldots, \mathcal{A}_n)$ be an awareness structure such that

- $\pi'(s, p) = \pi(s, p)$ if $\pi(s, p) \neq 1/2$ (the definition of π' if $\pi(s, p) = 1/2$ is irrelevant);

- $\mathcal{K}_i'(s) = \mathcal{K}_i(s)$ if $\mathcal{K}_i$ does not satisfy Generalized Reflexivity, and $\mathcal{K}_i'(s) = \mathcal{K}_i(s) \cup \{s\}$ otherwise;
- if $\emptyset \neq \mathcal{K}_i(s) \subseteq S_\Psi$ or if $\mathcal{K}_i(s) = \emptyset$ and $s \in S_\Psi$, then $\mathcal{A}_i(s)$ is the smallest set of formulas containing Ψ that is propositionally generated.

By construction, $M' \in \mathcal{N}_n^{C,pg}(\Phi)$. It is easy to check that if $C \cap \{t, e\} \neq \emptyset$, then agents know what they are aware of, so that $M' \in \mathcal{N}_n^{C,pd}(\Phi)$.

We complete the proof of part (a) by proving, by induction on the structure of φ, that if $s \in S_\Psi$ and $\Phi_\varphi \subseteq \Psi$, then $(M, s) \models \varphi$ iff $(M', s) \models \varphi_X$. If φ is either a primitive proposition, or $\varphi = \neg\psi$, or $\varphi = \varphi_1 \wedge \varphi_2$, the result is obvious either from the definition of π' or from the induction hypothesis. We omit details here.

Suppose that $\varphi = K_i\psi$. If $(M, s) \models K_i\psi$, then for all $t \in \mathcal{K}_i(s)$, $(M, t) \models \psi$. By the induction hypothesis, it follows that for all $t \in \mathcal{K}_i(s)$, $(M', t) \models \psi$. If $\mathcal{K}_i$ satisfies generalized reflexivity, it easily follows from Lemma C.1 that $(M, s) \models \psi$ so, by the induction hypothesis, $(M', s) \models \psi$. Hence, for all $t \in \mathcal{K}_i'(s)$, $(M', t) \models \psi$, so $(M', s) \models K_i\psi$. To show that $(M', s) \models X_i\psi$, it remains to show that $(M', s) \models A_i\psi$, that is, that $\psi \in \mathcal{A}_i(s)$. First suppose that $\emptyset \neq \mathcal{K}_i(s) \subseteq S_\Psi$. Since $(M, t) \models \psi$ for all $t \in \mathcal{K}_i(s)$, it follows that ψ is defined at all states in $\mathcal{K}_i(s)$. Thus, $\Phi_\psi \subseteq \Psi$, for otherwise a simple induction shows that ψ would be undefined at states in S_Ψ. Hence, $\Phi_\psi \subseteq \mathcal{A}_i(s)$. Since awareness is generated by primitive propositions, we have $\psi \in \mathcal{A}_i(s)$, as desired. Now suppose that $\mathcal{K}_i(s) = \emptyset$ and $s \in S_\Psi$. By assumption, $\Phi_\varphi = \Phi_\psi \subseteq \Psi \subseteq \mathcal{A}_i(s)$, so again $\psi \in \mathcal{A}_i(s)$.

For the converse, if $(M's) \models X_i\psi$, then $(M', s) \models K_i\psi$ and $\psi \in \mathcal{A}_i(s)$. By the definition of $\mathcal{A}_i$, $\mathcal{K}_i(s) \subseteq S_\Psi$, where $\Phi_\psi \subseteq \Psi$. Since $(M', s) \models K_i\psi$, $(M', t) \models \psi$ for all $t \in \mathcal{K}_i'(s)$. Therefore, by the induction hypothesis, $(M, t) \models \psi$ for all $t \in \mathcal{K}_i(s)$, which implies $(M, s) \models K_i\psi$, as desired.

For part (b), given $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi, \mathcal{A}_1, \ldots, \mathcal{A}_n) \in \mathcal{N}_n^{C,pd}(\Phi)$, let $M' = (\Sigma', \mathcal{K}_1', \ldots, \mathcal{K}_n', \pi'$, $\{\rho_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\})$ be an HMS structure such that

- $\Sigma' = \Sigma \times 2^{\Phi}$;
- $S_{\Psi} = \Sigma \times \{\Psi\}$ for $\Psi \subseteq \Phi$;
- $\pi'((s,\Psi),p) = \pi(s,p)$ if $p \in \Psi$ and $\pi'((s,\Psi),p) = 1/2$ otherwise;
- $\mathcal{K}_i'((s,\Psi)) = \{(t, \Psi \cap \Psi_i(s)) : t \in \mathcal{K}_i(s)\}$, where $\Psi_i(s) = \{p : p \in \mathcal{A}_i(s)\}$ is the set of primitive propositions that agent i is aware of at state s;
- $\rho_{\Psi',\Psi}((s,\Psi')) = (s,\Psi)$.

Note that since agents know what they are aware of, if $t \in \mathcal{K}_i(s)$, then $\Psi_i(t) = \Psi_i(s)$.

We first show that M' satisfies confinedness and that projections preserve knowledge and ignorance. Confinedness follows since $\mathcal{K}_i'((s,\Psi)) \subseteq S_{\Psi \cap \Psi_i(s)}$. To prove projections preserve knowledge, suppose that $\Psi_1 \subseteq \Psi_2 \subseteq \Psi_3$ and $\mathcal{K}_i((s,\Psi_3)) \subseteq S_{\Psi_2}$. Then $\Psi_2 = \Psi_3 \cap \Psi_i(s)$ and $\Psi_1 = \Psi_1 \cap \Psi_i(s)$. Thus $\rho_{\Psi_2,\Psi_1}(\mathcal{K}_i'((s,\Psi_3))) = \{(t,\Psi_1) : (t, \Psi_3 \cap \Psi_i(s)) \in \mathcal{K}_i'((s,\Psi_3))\} = \{(t,\Psi_1) : t \in \mathcal{K}_i(s)\}$. Similarly, $\mathcal{K}_i'(\rho_{\Psi_3,\Psi_1}(s,\Psi_3)) = \{(t, \Psi_1 \cap \Psi_i(s)) : t \in \mathcal{K}_i(s)\}$. Therefore, projections preserve knowledge.

To prove that projections preserve ignorance, note that $\mathcal{K}_i'(\rho_{\Psi',\Psi}(s,\Psi')) = \mathcal{K}_i'((s,\Psi)) = \{(t, \Psi \cap \Psi_i(s)) : t \in \mathcal{K}_i(s)\}$ and $\mathcal{K}_i'((s,\Psi')) = \{(t, \Psi' \cap \Psi_i(s)) : t \in \mathcal{K}_i(s)\}$. If $(s,\Psi'') \in (\mathcal{K}_i'((s,\Psi')))^{\uparrow}$, then $\Psi' \cap \Psi_i(s) \subseteq \Psi''$. Since $\Psi \subseteq \Psi'$, it follows that $\Psi \cap \Psi_i(s) \subseteq \Psi''$. Hence $(s,\Psi'') \in (\mathcal{K}_i'(\rho_{\Psi',\Psi}(s,\Psi')))^{\uparrow}$. Therefore, projection preserves ignorance.

We now show by induction on the structure of φ that if $\Phi_\varphi \subseteq \Psi$, then $(M,s) \models \varphi_X$ iff $(M',(s,\Psi)) \models \varphi$. If φ is a primitive proposition, or $\varphi = \neg\psi$, or $\varphi = \varphi_1 \wedge \varphi_2$, the result is obvious either from the definition of π' or from the induction hypothesis. We omit details here.

Suppose that $\varphi = K_i\psi$. If $(M',(s,\Psi)) \models K_i\psi$, then for all $(t, \Psi \cap \Psi_i(s)) \in \mathcal{K}_i'((s,\Psi))$, $(M',(t,\Psi \cap \Psi_i(s))) \models \psi$. By the induction hypothesis and the definition of $\mathcal{K}_i'$, it follows that for all $t \in \mathcal{K}_i(s)$, $(M,t) \models \psi$, so $(M,s) \models K_i\psi$. Also note that if $(M',(t,\Psi \cap \Psi_i(s))) \models \psi$, then ψ is defined at all states in

$S_{\Psi\cap\Psi_i(s)}$, and therefore at all states in $S_{\Psi_i(s)}$. Hence, $\Phi_\psi \subseteq \Psi_i(s) \subseteq \mathcal{A}_i(s)$. Since awareness is generated by primitive propositions, $\psi \in \mathcal{A}_i(s)$. Thus, $(M,s) \models A_i\psi$, which implies that $(M,s) \models X_i\psi$, as desired.

For the converse, suppose that $(M,s) \models X_i\psi$ and $\Phi_\varphi \subseteq \Psi$. Then $(M,s) \models K_i\psi$ and $\psi \in \mathcal{A}_i(s)$. Since $\psi \in \mathcal{A}_i(s)$, $\Phi_\psi \subseteq \Psi_i(s)$. Hence, $\Phi_\psi \subseteq (\Psi \cap \Psi_i(s))$. $(M,s) \models K_i\psi$ implies that $(M,t) \models \psi$ for all $t \in \mathcal{K}_i(s)$. By the induction hypothesis, since $\Phi_\psi \subseteq \Psi$, $(M',(t,\Psi)) \models \psi$ for all $t \in \mathcal{K}_i(s)$. Since $\Phi_\psi \subseteq (\Psi \cap \Psi_i(s))$, by Lemma C.1, it follows that $(M',(t,\Psi\cap\Psi_i(s))) \models \psi$ for all $(t,\Psi\cap\Psi_i(s)) \in \mathcal{K}_i'((s,\Psi))$. Thus, $(M',(s,\Psi)) \models K_i\psi$, as desired.

We now show that $M' \in \mathcal{H}_n^C(\Phi)$. If $\mathcal{K}_i$ is reflexive, then $(s,\Psi\cap\Psi_i(s)) \in \mathcal{K}_i'((s,\Psi))$, so $(s,\Psi) \in (\mathcal{K}_i'((s,\Psi)))^\uparrow$. Thus, M' satisfies generalized reflexivity. Now suppose that $\mathcal{K}_i$ is transitive. If $(s',\Psi\cap\Psi_i(s)) \in \mathcal{K}_i'((s,\Psi))$, then $s' \in \mathcal{K}_i(s)$ and $\mathcal{K}_i'((s',\Psi\cap\Psi_i(s))) = \{(t,\Psi\cap\Psi_i(s)\cap\Psi_i(s')) : t \in \mathcal{K}_i(s')\}$. Since agents know what they are aware of, if $s' \in \mathcal{K}_i(s)$, then $\Psi_i(s) = \Psi_i(s')$, so $\mathcal{K}_i'((s',\Psi\cap\Psi_i(s))) = \{(t,\Psi\cap\Psi_i(s)) : t \in \mathcal{K}_i(s')\}$. Since $\mathcal{K}_i$ is transitive, $\mathcal{K}_i(s') \subseteq \mathcal{K}_i(s)$, so $\mathcal{K}_i'((s',\Psi\cap\Psi_i(s))) \subseteq \mathcal{K}_i'((s,\Psi))$. Thus, M' satisfies part (a) of stationarity. Finally, suppose that $\mathcal{K}_i$ is Euclidean. If $(s',\Psi\cap\Psi_i(s)) \in \mathcal{K}_i'((s,\Psi))$, then $s' \in \mathcal{K}_i(s)$ and $\mathcal{K}_i'((s',\Psi\cap\Psi_i(s))) = \{(t,\Psi\cap\Psi_i(s)\cap\Psi_i(s')) : t \in \mathcal{K}_i(s')\}$. Since agents know what they are aware of, if $s' \in \mathcal{K}_i(s)$, then $\Psi_i(s) = \Psi_i(s')$, so $\mathcal{K}_i'((s',\Psi\cap\Psi_i(s))) = \{(t,\Psi\cap\Psi_i(s)) : t \in \mathcal{K}_i(s')\}$. Since $\mathcal{K}_i$ is Euclidean, $\mathcal{K}_i(s') \supseteq \mathcal{K}_i(s)$, so $\mathcal{K}_i'((s',\Psi\cap\Psi_i(s))) \supseteq \mathcal{K}_i'((s,\Psi))$. Thus M' satisfies part (b) of stationarity.

If $\{t,e\} \cap C = \emptyset$, then it is easy to check that the result holds even if $M \in (\mathcal{N}_n^{C,pg}(\Phi) - \mathcal{N}_n^{C,pd}(\Phi))$, since the property that agents know what they are aware of was only used to prove part (a) and part (b) of stationarity in the proof. ∎

Corollary 3.3: *If $C \subseteq \{r,t,e\}$ then*

(a) if $C \cap \{t,e\} = \emptyset$, then φ is weakly valid in $\mathcal{H}_n^C(\Phi)$ iff φ_X is valid in $\mathcal{N}_n^{C,pg}(\Phi)$.

(b) if $C \cap \{t,e\} \neq \emptyset$, then φ is weakly valid in $\mathcal{H}_n^C(\Phi)$ iff φ_X is valid in $\mathcal{N}_n^{C,pd}(\Phi)$.

Proof: For part (a), suppose that φ_X is valid with respect to the class of awareness structures $\mathcal{N}_n^{C,pg}(\Phi)$ where awareness is generated by primitive propositions and that φ is not weakly valid with respect to the class of HMS structures $\mathcal{H}_n^C(\Phi)$. Then $\neg\varphi$ is true at some state in some HMS structure in $\mathcal{H}_n^C(\Phi)$. By part (a) of Theorem 3.2, $\neg\varphi_X$ is also true at some state in some awareness structure where awareness is generated by primitive propositions, a contradiction since φ_X is valid in $\mathcal{N}_n^{C,pg}(\Phi)$.

For the converse, suppose that φ is weakly valid in $\mathcal{H}_n^C(\Phi)$ and that φ_X is not valid with respect to the class of awareness structures $\mathcal{N}_n^{C,pg}(\Phi)$. Then $\neg\varphi_X$ is true at some state in some awareness structure in $\mathcal{N}_n^{C,pg}(\Phi)$. By part (b) of Theorem 3.2, $\neg\varphi$ is also true at some state in some HMS structure in $\mathcal{H}_n^C(\Phi)$, a contradiction since φ is weakly valid in $\mathcal{H}_n^C(\Phi)$.

The proof of part (b) is the same except that $\mathcal{N}_n^{C,pg}(\Phi)$ is replaced throughout by $\mathcal{N}_n^{C,pd}(\Phi)$. ∎

Theorem 4.2: *Let $\mathcal{C}$ be a (possibly empty) subset of $\{\mathrm{T}', 4', 5'\}$ and let C be the corresponding subset of $\{r, t, e\}$. Then $\mathrm{AX}_n^{K,\hookrightarrow} \cup \mathcal{C}$ is a sound and complete axiomatization of the language $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ with respect to $\mathcal{H}_n^C(\Phi)$.*

Proof: Soundness is straightforward, as usual, by induction on the length of the proof (after showing that all the axioms are sound and that the inference rules preserve strong validity). We leave details to the reader.

To prove completeness, we first define a *simplified HMS* structure for n agents to be a tuple $M = (\Sigma, \mathcal{K}_1, \ldots, \mathcal{K}_n, \pi)$. That is, a simplified HMS structure is an HMS structure without the projection functions. The definition of $\models$ for simplified HMS structures is the same as that for HMS structures. (Recall that the projections functions are not needed for defining $\models$.) Let $\mathcal{H}_n^-(\Phi)$ consist of all simplified HMS structures for n agents over Φ that satisfy confinedness.

Lemma C.2: $\mathrm{AX}_n^{K,\hookrightarrow}$ *is a sound and complete axiomatization of the language $\mathcal{L}_n^{K,\hookrightarrow}(\Phi)$ with respect to $\mathcal{H}_n^-(\Phi)$.*

Proof: Again, soundness is obvious. For completeness, it clearly suffices to show that every $\mathrm{AX}_n^{K,\hookrightarrow}$-

consistent formula is satisfiable in some structure in $\mathcal{H}_n^-$. As usual, we do this by constructing a canonical model $M^c \in \mathcal{H}_n^-$ and showing that every $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent formula of the form $\varphi = i$ for $i \in \{0, 1/2, 1\}$ is satisfiable in some state of M^c. Since $(M^c, s) \models \varphi$ iff $(M^c, s) \models \varphi = 1$, this clearly suffices to prove the result.

Let $M^c = (\Sigma^c, \mathcal{K}_1^c, ..., \mathcal{K}_n^c, \pi^c)$, where

- $S_\Psi^c = \{s_V : V$ is a maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent subset D_2 and for all $p \in (\Phi - \Psi),\ p = 1/2 \in V$, and for all $p \in \Psi$, $(p = 1/2) \notin V\}$;
- $\Sigma^c = \cup_{\Psi \subseteq \Phi} S_\Psi^c$;
- $\mathcal{K}_i^c(s_V) = \{s_W : V/K_i \subseteq W\}$, where $V/K_i = \{\varphi = 1 : K_i\varphi = 1 \in V\}$;
- $\pi^c(s_V, p) = \begin{cases} 1 & \text{if } p = 1 \in V \\ 0 & \text{if } p = 0 \in V \\ 1/2 & \text{if } p = 1/2 \in V. \end{cases}$

Note that, by Lemma B.4(2), the interpretation π^c is well defined.

We want to show that

$$(M^c, s_V) \models \psi = j \ \text{ iff } \ \psi = j \in V, \text{for } j \in \{0, 1/2, 1\}. \tag{1}$$

We show that (1) holds by induction on the structure of formulas. If ψ is a primitive proposition, this follows from the definition of $\pi^c(s_V, \psi)$. If $\psi = \neg\varphi$ or $\psi = \varphi_1 \wedge \varphi_2$ or $\psi = \varphi_1 \hookrightarrow \varphi_2$, the argument is similar to that of Lemma B.5, we omit details here.

If $\psi = K_i\varphi$, then by the definition of V/K_i, if $\psi = 1 \in V$, then $\varphi = 1 \in V/K_i$, which implies that if $s_W \in \mathcal{K}_i^c(s_V)$, then $\varphi = 1 \in W$. Moreover, by axiom B1, $\varphi = 1/2 \notin V$. By the induction hypothesis, this implies that $(M^c, s_V) \models \neg(\varphi = 1/2)$ and that $(M^c, s_W) \models \varphi$ for all W such that $s_W \in \mathcal{K}_i^c(s_V)$. This in turn implies that $(M^c, s_V) \models K_i\varphi$. Thus, $(M^c, s_V) \models (K_i\varphi) = 1$, i.e., $(M^c, s_V) \models \psi = 1$.

For the other direction, the argument is essentially identical to analogous arguments for Kripke structures. Suppose that $(M^c, s_V) \models (K_i\varphi) = 1$. It follows that the set $(V/K_i) \cup \{\neg(\varphi = 1)\}$ is not $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. For suppose otherwise. By Lemma B.3, there would be a maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent set W that contains $(V/K_i) \cup \{\neg(\varphi = 1)\}$ and, by construction, we would have $s_W \in \mathcal{K}_i^c(s_V)$. By the induction hypothesis, $(M^c, s_W) \not\models (\varphi = 1)$, and so $(M^c, s_W) \not\models \varphi$. Thus, $(M^c, s_V) \not\models K_i\varphi$, contradicting our assumption. Since $(V/K_i) \cup \{\neg(\varphi = 1)\}$ is not $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent, there must be some finite subset, say $\{\varphi_1, ..., \varphi_k, \neg(\varphi = 1)\}$, which is not $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. By classical propositional reasoning (which can be applied since all formulas are in D_2),

$$\mathrm{AX}_n^{K,\hookrightarrow} \vdash \varphi_1 \hookrightarrow (\varphi_2 \hookrightarrow (... \hookrightarrow (\varphi_k \hookrightarrow (\varphi = 1)))).$$

By Gen,

$$\mathrm{AX}_n^{K,\hookrightarrow} \vdash K_i(\varphi_1 \hookrightarrow (\varphi_2 \hookrightarrow (... \hookrightarrow (\varphi_k \hookrightarrow (\varphi = 1))))). \tag{2}$$

Using axiom K′ and classical propositional reasoning, we can show by induction on k that

$$\begin{aligned} \mathrm{AX}_n^{K,\hookrightarrow} \quad \vdash \quad & K_i(\varphi_1 \hookrightarrow (\varphi_2 \hookrightarrow (... \hookrightarrow (\varphi_k \hookrightarrow (\varphi = 1))))) \hookrightarrow \\ & \hookrightarrow (K_i\varphi_1 \hookrightarrow (K_i\varphi_2 \hookrightarrow (... \hookrightarrow (K_i\varphi_k \hookrightarrow (K_i(\varphi = 1)))))). \end{aligned} \tag{3}$$

Now by MP′ and Equations (2) and (3), we get

$$\mathrm{AX}_n^{K,\hookrightarrow} \vdash (K_i\varphi_1 \hookrightarrow (K_i\varphi_2 \hookrightarrow (... \hookrightarrow (K_i\varphi_k \hookrightarrow (K_i(\varphi = 1)))))).$$

By Lemma B.4(10), it follows that

$$K_i\varphi_1 \hookrightarrow (K_i\varphi_2 \hookrightarrow (... \hookrightarrow (K_i\varphi_k \hookrightarrow (K_i(\varphi = 1))))) \in V.$$

Since $\varphi_1, ..., \varphi_k \in V/K_i$, there exist formulas $\alpha_1, \ldots, \alpha_k$ such that φ_i has the form $\alpha_i = 1$, for $i = 1, \ldots, k$. By definition of V/K_i, $(K_i\alpha_1) = 1, ..., (K_i\alpha_k) = 1 \in V$. Note that, by Prop′, $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \alpha_j \hookrightarrow (\alpha_j = 1)$. So, by Gen, K′, Prop′, and MP′, $\mathrm{AX}_n^{K,\hookrightarrow} \vdash K_i\alpha_j \hookrightarrow K_i(\alpha_j = 1)$. Thus, $K_i\alpha_j \hookrightarrow K_i(\alpha_j = 1) \in V$.

Another application of Prop$'$ gives that $\mathrm{AX}_n^{K,\hookrightarrow} \vdash ((K_i\alpha_j) = 1 \wedge (K_i\alpha_j \hookrightarrow K_i(\alpha_j = 1)) \hookrightarrow K_i(\alpha_j = 1)$. Since $(K_i\alpha_j) = 1 \wedge (K_i\alpha_j \hookrightarrow K_i(\alpha_j = 1) \in V$, it follows that $K_i(\alpha_j = 1) \in V$; i.e., $K_i\varphi_j \in V$. Note that, by Prop$'$, $\mathrm{AX}_n^{K,\hookrightarrow} \vdash (\beta \wedge (\beta \hookrightarrow \gamma)) \hookrightarrow \gamma$. By repeatedly applying this observation and Lemma B.4(3), we get that $K_i(\varphi = 1) \in V$. Since, $(M^c, s_V) \models (K_i\varphi) = 1$ implies $(M^c, s_V) \not\models \varphi = 1/2$, it follows by the induction hypothesis that $\varphi = 1/2 \notin V$. Therefore $(\varphi = 0 \vee \varphi = 1) \in V$, so by axiom B2 and Lemma B.4(3), $(K_i\varphi) = 1 \in V$, as desired.[13]

Finally, by axiom B1 and Lemma B.4(3), $(K_i\varphi) = 1/2 \in V$ iff $\varphi = 1/2 \in V$. By the induction hypothesis, $\varphi = 1/2 \in V$ iff $(M^c, s_V) \models \varphi = 1/2$. By the definition of $\models$, $(M^c, s_V) \models \varphi = 1/2$ iff $(M^c, s_V) \models K_i\varphi = 1/2$.

This completes the proof of (1). Since every $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent formula φ is in some maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent set, φ must be satisfied at some state in M^c.

It remains to show that M^c satisfies confinedness. So suppose that $s_V \in S_\Psi$. We must show that $\mathcal{K}_i^c(s_V) \subseteq S_{\Psi'}$ for some $\Psi' \subseteq \Psi$. This is equivalent to showing that, for all $s_W, s_{W'} \in \mathcal{K}_i^c(s_V)$ and all primitive propositions p, (a) $(M^c, s_W) \models p = 1/2$ iff $(M^c, s_{W'}) \models p = 1/2$ and (b) if $(M^c, s_V) \models p = 1/2$, then $(M^c, s_W) \models p = 1/2$. For (a), suppose that $s_W, s_{W'} \in \mathcal{K}_i(s)$ and $(M^c, s_W) \models p = 1/2$. Since $s_W \in \mathcal{K}_i^c(s_V)$, we must have $(M^c, s_V) \models \neg K_i((p \vee \neg p) = 1)$. V contains every instance of Conf2. Thus, by (1), $(M^c, s_V) \models \neg K_i(p = 1/2) \hookrightarrow K_i((p \vee \neg p) = 1)$. It follows that $(M^c, s_V) \models K_i(p = 1/2)$. Thus, $(M^c, s_{W'}) \models p = 1/2$, as desired. For (b), suppose that $(M^c, s_V) \models p = 1/2$. Since V contains every instance of Conf1, it follows from (1) that $(M^c, s_V) \models p = 1/2 \hookrightarrow K_i(p = 1/2)$. It easily follows that $(M^c, s_W) \models p = 1/2$. Thus, $M^c \in \mathcal{H}_n^-$, as desired.

To finish the proof that $\mathrm{AX}_n^{K,\hookrightarrow}$ is complete with respect to $\mathcal{H}_n^-(\Phi)$, suppose that φ is valid in $\mathcal{H}_n^-(\Phi)$. This implies that $(\varphi = 0) \vee (\varphi = 1/2)$ is not satisfiable, so by (1), $(\varphi = 0) \vee (\varphi = 1/2)$ is not $\mathrm{AX}_n^{K,\hookrightarrow}$-

[13]This proof is almost identical to the standard modal logic proof that $(M, s_V) \models K_i\varphi$ implies $K_i\varphi \in V$ [Fagin, Halpern, Moses, and Vardi 1995].

consistent. Thus, $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \neg((\varphi = 0) \vee (\varphi = 1/2))$. By Prop′ and MP′, it follows that $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \varphi = 1$ and $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \varphi$, as desired. ∎

We now want to show that there exist projection functions $\rho^c_{\Psi',\Psi}$ such that $(\Sigma^c, \mathcal{K}^c_1, \ldots, \mathcal{K}^c_n, \pi^c, \{\rho^c_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\}) \in \mathcal{H}_n(\Phi)$. The intention is to define $\rho^c_{\Psi',\Psi}$ so that $\rho^c_{\Psi',\Psi}(s_V) = s_W$, where $s_W \in S^c_\Psi$ and agrees with s_V on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi)$. (We say that s_V agrees with s_W on φ if $(M^c, s_V) \models \varphi$ iff $(M^c, s_W) \models \varphi$.) But first we must show that this is well-defined; that is, that there exists a unique W with these properties. To this end, let $R_{\Psi',\Psi}$ be a binary relation on states in Σ^c such that $R_{\Psi',\Psi}(s_V, s_W)$ holds if $s_V \in S^c_{\Psi'}$, $s_W \in S^c_\Psi$, and s_V and s_W agree on formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi)$. We want to show that $R_{\Psi',\Psi}$ actually defines a function; that is, for each state $s_V \in S^c_{\Psi'}$, there exists a unique $s_W \in S^c_\Psi$ such that $R_{\Psi',\Psi}(s_V, s_W)$. The following lemma proves existence.

Lemma C.3: *If $\Psi \subseteq \Psi'$, then for all $s_V \in S^c_{\Psi'}$, there exists $s_W \in S^c_\Psi$ such that $R_{\Psi',\Psi}(s_V, s_W)$ holds.*

Proof: Suppose that $s_V \in S^c_{\Psi'}$. Let V_Ψ be the subset of V containing all formulas of the form $\varphi = 1$, where φ contains only primitive propositions in Ψ. It is easily seen that $V_\Psi \cup \{p = 1/2 : p \notin \Psi\}$ is $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. For suppose, by way of contradiction, that it is not $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. So, without loss of generality, there exists a formula ψ such that $\psi = 1 \in V_\Psi$ and $\mathrm{AX}_n^{K,\hookrightarrow} \vdash p_1 = 1/2 \hookrightarrow (p_2 = 1/2 \hookrightarrow (\ldots \hookrightarrow (p_k = 1/2 \hookrightarrow \neg(\psi = 1))))$, where $p_i \neq p_j$ for $i \neq j$ and $p_1, \ldots, p_k \in (\Phi - \Psi)$. By Lemma C.2, it follows that $\mathcal{H}_n^- \models p_1 = 1/2 \hookrightarrow (p_2 = 1/2 \hookrightarrow (\ldots \hookrightarrow (p_k = 1/2 \hookrightarrow \neg(\psi = 1))))$. It easily follows that $\mathcal{H}_n^- \models \neg(\psi = 1)$. Applying Lemma C.2 again, we get that $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \neg(\psi = 1)$. This is a contradiction, since $\psi = 1 \in V$ and V is a maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent subset of D_2. It follows that $V_\Psi \cup \{p = 1/2 : p \notin \Psi\}$ is contained in some maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent subset W of D_2. So s_V and s_W agrees on all formulas of the form $\varphi = 1$ for $\varphi \in \mathcal{L}_n^{K,\hookrightarrow}(\Psi)$ and therefore agree on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi)$, i.e., $R_{\Psi',\Psi}(s_V, s_W)$ holds. ∎

The next lemma proves uniqueness.

Lemma C.4: *If $\Psi \subseteq \Psi'$, then for all $s_V \in S^c_{\Psi'}$, $s_W, s_{W'} \in S^c_{\Psi}$, if $R_{\Psi',\Psi}(s_V, s_W)$ and $R_{\Psi',\Psi}(s_V, s_{W'})$ both hold, then $W = W'$.*

Proof: Suppose that $R_{\Psi',\Psi}(s_V, s_W)$ and $R_{\Psi',\Psi}(s_V, s_{W'})$ both hold. We want to show that $W = W'$.

Define a formula ψ to be *simple* if it is a Boolean combination of formulas of the form $\varphi = k$, where φ is implication-free. It is easy to check that if φ is implication-free, since $s_W \in S_\Psi$, then $(M^c, s_W) \models \varphi = 1/2$ iff $\Phi_\varphi - \Psi \neq \emptyset$; the same is true for $s_{W'}$. Moreover, if $\Phi_\varphi \subseteq \Psi$, then s_W and $s_{W'}$ agree on φ. Thus, it easily follows that s_W and $s_{W'}$ agree on all simple formulas. We show that $W = W'$ by showing that every formula is equivalent to a simple formula; that is, for every formula $\varphi \in D_2$, there exists a simple formula φ' such that $\mathcal{H}_n^- \models \varphi \rightleftharpoons \varphi'$.

First, we prove this for formulas φ of the form $\psi = k$, by induction on the structure of ψ. If ψ is a primitive proposition p, then φ is simple. The argument is straightforward, using the semantic definitions, if ψ is of the form $\neg\psi'$, $\psi_1 \wedge \psi_2$, or $\psi_1 \hookrightarrow \psi_2$.

If ψ has the form $K_i\psi'$, we proceed by cases. If $k = 1/2$, then the result follows immediately from the induction hypothesis, using the observation that $\mathcal{H}_n^- \models K_i\psi' = 1/2 \rightleftharpoons \psi' = 1/2$. To deal with the case $k = 1$, for $\Phi' \subseteq \Phi_\psi$, define $\sigma_{\psi,\Phi'} = \bigwedge_{p\in\Phi'}((p \vee \neg p) = 1) \wedge \bigwedge_{p\in(\Phi_\psi - \Phi')} p = 1/2$. By the induction hypothesis, $\psi' = 1$ is equivalent to a simple formula ψ''. Moreover, ψ'' is equivalent to $\bigvee_{\Phi'\subseteq\Phi_\psi}(\psi'' \wedge \sigma_{\psi,\Phi'})$. Finally, note that $\psi'' \wedge \sigma_{\psi,\Phi'}$ is equivalent to a formula where each subformula $\xi = k$ of ψ'' such that $\Phi_\xi - \Phi' \neq \emptyset$ is replaced by $\top$ if $k = 1/2$ and replaced by $\perp$ if $k \neq 1/2$; each subformula of the form $\xi = 1/2$ such that $\Phi_\xi \subseteq \Phi'$ is replaced by $\perp$. Thus, $\psi'' \wedge \sigma_{\psi,\Phi'}$ is equivalent to a formula of the form $\psi_{\Phi'} \wedge \sigma_{\psi,\Phi'}$, where $\psi_{\Phi'}$ is simple, all of its primitive propositions are in Φ', and all of its subformulas have the form $\xi = 0$ or $\xi = 1$. Let $\sigma^+_{\psi,\Phi'} = \bigwedge_{p\in\Phi'}((p \vee \neg p) = 1)$ and let $\sigma^-_{\psi,\Phi'} = \bigwedge_{p\in(\Phi_\psi - \Phi')}(p = 1/2)$ (so that $\sigma_{\psi,\Phi'} = \sigma^+_{\psi,\Phi'} \wedge \sigma^-_{\psi,\Phi'}$). An easy induction on the structure of a formula shows that $\psi_{\Phi'} \wedge \sigma^+_{\psi,\Phi'}$ is equivalent to $\xi' = 1 \wedge \sigma^+_{\psi,\Phi'}$, where ξ' is an implication-free formula. Finally, it is easy to see that $\xi' = 1 \wedge \sigma^+_{\psi,\Phi'}$ is equivalent to a formula $\xi''_{\Phi'} = 1$, where $\xi''_{\Phi'}$ is implication-free. To summarize, we have

$$\mathcal{H}_n^- \models \psi' = 1 \rightleftharpoons \bigvee_{\Phi' \subseteq \Phi_\psi} (\xi''_{\Phi'} = 1 \wedge \sigma^-_{\psi,\Phi'}).$$

It easily follows that we have

$$\mathcal{H}_n^- \models K_i(\psi' = 1) \rightleftharpoons K_i \left(\bigvee_{\Phi' \subseteq \Phi} (\xi''_{\Phi'} = 1 \wedge \sigma^-_{\psi,\Phi'}) \right). \tag{4}$$

It follows from confinedness that

$$\mathcal{H}_n^- \models K_i \left(\bigvee_{\Phi' \subseteq \Phi} (\xi''_{\Phi'} = 1 \wedge \sigma^-_{\psi,\Phi'}) \right) \rightleftharpoons \bigvee_{\Phi' \subseteq \Phi} K_i(\xi''_{\Phi'} = 1 \wedge \sigma^-_{\psi,\Phi'}). \tag{5}$$

Since $\mathcal{H}_n^- \models K_i(\psi_1 \wedge \psi_2) \rightleftharpoons K_i\psi_1 \wedge K_i\psi_2$ and $\mathcal{H}_n^- \models \xi = 1/2 \rightleftharpoons K_i(\xi = 1/2)$, it follows that

$$\mathcal{H}_n^- \models K_i(\xi''_{\Phi'} = 1 \wedge \sigma^-_{\psi,\Phi'}) \rightleftharpoons K_i(\xi''_{\Phi'} = 1) \wedge \sigma^-_{\psi,\Phi'}. \tag{6}$$

Finally, since $\mathcal{H}_n^- \models K_i(\xi = 1) \rightleftharpoons K_i\xi = 1$, we can conclude from (4), (5), and (6) that

$$\mathcal{H}_n^- \models (K_i\psi') = 1 \rightleftharpoons (K_i\xi''_{\Phi'}) = 1 \wedge \sigma^-_{\psi,\Phi'},$$

and hence $K_i\psi' = 1$ is equivalent to a simple formula.

Since $K_i\psi = 0$ is equivalent to $\neg(K_i\psi = 1) \wedge \neg(K_i\psi = 1/2)$, and each of $K_i\psi = 1$ and $K_i\psi = 1/2$ is equivalent to a simple formula, it follows that $K_i\psi = 0$ is equivalent to a simple formula.

The arguments that $\neg\psi_1$, $\psi_1 \wedge \psi_2$, $\psi_1 \hookrightarrow \psi_2$, and $K_i\psi_1$ are equivalent to simple formulas if ψ_1 and ψ_2 are follows similar lines, and is left to the reader. It follows that every formula in D_2 is equivalent to a simple formula. This shows that $W = W'$, as desired. ∎

It follows from Lemmas C.3 and C.4 that $R_{\Psi,\Psi'}$ defines a function. We take this to be the definition of $\rho^c_{\Psi,\Psi'}$. We now must show that the projection functions are coherent.

Lemma C.5: *If* $\Psi_1 \subseteq \Psi_2 \subseteq \Psi_3$*, then* $\rho^c_{\Psi_3,\Psi_1} = \rho^c_{\Psi_3,\Psi_2} \circ \rho^c_{\Psi_2,\Psi_1}$.

Proof: If $s_V \in S_{\Psi_3}$, we must show that $\rho^c_{\Psi_3,\Psi_1}(s_V) = \rho^c_{\Psi_2,\Psi_1}(\rho^c_{\Psi_3,\Psi_2}(s_V))$. Let $s_W = \rho^c_{\Psi_3,\Psi_2}(s_V)$ and $s_X = \rho^c_{\Psi_2,\Psi_1}(s_W)$. Then s_W and s_V agree on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi_2)$ and s_W and s_X agree on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi_1)$. Thus, s_V and s_X agree on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi_1)$. Moreover, by construction, $s_X \in S_{\Psi_1}$. By Lemma C.4, we must have $s_X = \rho^c_{\Psi_3,\Psi_1}(s_V)$. ∎

Since we have now shown that the projection functions are well defined, from here on, we abuse notation and refer to M^c as $(\Sigma^c, \mathcal{K}_1^c, \ldots, \mathcal{K}_n^c, \pi^c, \{\rho^c_{\Psi',\Psi} : \Psi \subseteq \Psi' \subseteq \Phi\})$. To complete the proof of Theorem 4.2, we now show that M^c satisfies projection preserves knowledge and ignorance. Both facts will follow easily from Proposition C.7 below.

We first need a lemma, which provides a condition for s_W to be in $\mathcal{K}_i^c(S_V)$ that is easier to check.

Lemma C.6: *If $\mathcal{K}_i^c(s_V) \subseteq S^c_\Psi$, $s_W \in S^c_\Psi$, and $V/K_i \cap \mathcal{L}_n^{K,\hookrightarrow}(\Psi) \subseteq W$, then $s_W \in \mathcal{K}_i^c(s_V)$.*

Proof: Suppose that V and W are as in the antecedent of the statement of the lemma. We must show that $V/K_i \subseteq W$.

First note that V/K_i is closed under implication. That is, if $\varphi_1 = 1 \in V/K_i$, and $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \varphi_1 \hookrightarrow \varphi_2$, then $\varphi_2 = 1 \in V/K_i$. This follows from the observations that if $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \varphi_1 \hookrightarrow \varphi_2$, then $\mathrm{AX}_n^{K,\hookrightarrow} \vdash K_i(\varphi_1 \hookrightarrow \varphi_2)$, and $\mathrm{AX}_n^{K,\hookrightarrow} \vdash K_i\varphi_1 \hookrightarrow K_i\varphi_2$; so, by Prop$'$, $\mathrm{AX}_n^{K,\hookrightarrow} \vdash (K_i\varphi_1 \hookrightarrow K_i\varphi_2) = 1$. Thus, $(K_i\varphi_1 \hookrightarrow K_i\varphi_2) = 1 \in V$. Moreover, since $\varphi_1 = 1 \in V/K_i$, we must have $K_i\varphi_1 = 1 \in V$. Finally, by Prop$'$, $\mathrm{AX}_n^{K,\hookrightarrow} \vdash (K_i\varphi_1 = 1 \wedge (K_i\varphi_1 \hookrightarrow K_i\varphi_2) = 1) \hookrightarrow K_i\varphi_2 = 1$. Thus, Lemma B.4(3), $K_i\varphi_2 = 1 \in V$. So $\varphi_2 = 1 \in W/K_i$, as desired. By Lemma B.4(3), W is also closed under implication. Thus, by the proof of Lemma C.4, it suffices to show that $\varphi = 1 \in W$ for each simple formula φ such that $K_i\varphi = 1 \in V$. Indeed, we can take φ to be in conjunctive normal form: a conjunction of disjunctions of formulas of *basic formulas*, that is formulas of the form $\psi = k$ where ψ is implication-free. Moreover, since by Lemma B.4(4), W is closed under conjunction ($\varphi_1 \in W$ and $\varphi_2 \in W$ implies that $\varphi_1 \wedge \varphi_2 \in W$) and it is easy to show that V/K_i is closed under breaking up of conjunctions (if $\varphi_1 \wedge \varphi_2 \in V/K_i$ then $\varphi_i \in V/K_i$ for $i = 1, 2$), it suffices to show that $\psi = 1 \in W$ for each disjunction ψ of basic formulas such that $K_i\psi = 1 \in V$. We proceed by induction on the number of disjuncts in ψ.

If there is only one disjunct in ψ, that is, ψ has the form $\psi' = k$, where ψ' is implication-free, suppose first that $k = 1/2$. It is easy to check that $\mathcal{H}_n^- \models (\psi' = 1/2) \rightleftharpoons \bigvee_{p\in\Phi_{\psi'}}(p = 1/2)$ and $\mathcal{H}_n^- \models K_i(\psi' = 1/2) \rightleftharpoons \bigvee_{p\in\Phi_{\psi'}} K_i(p = 1/2)$. Since $(K_i(\psi' = 1/2)) = 1 \in V$, by Prop$'$ and Lemma B.4(3) $K_i(\psi' = 1/2) \in V$ and $K_i(p = 1/2) \in V$ for some $p \in \Phi_{\psi'}$ (since, by Lemma C.2, for all formulas σ, we have $\sigma \in V$ iff $(M^c, s_V) \models \sigma$). Since $\mathcal{K}_i^c(s_V) \subseteq S^c_\Psi$, it must be the case that $K_i(p = 1/2) \in V$ iff $p \notin \Psi$. Since $s_W \in S^c_\Psi$, $p = 1/2 \in W$. Since W is closed under implication, $\psi' = 1/2 \in W$, as desired. If $k = 0$ or $k = 1$, then it is easy to see that $\psi = 1 \in V/K_i$ only if $\psi = 1 \in \mathcal{L}_n^{K,\hookrightarrow}(\Psi)$ so, by assumption, $\psi = 1 \in W$.

If ψ has more than one disjunct, suppose that there is some disjunct of the form $\psi' = 1/2$ in ψ. If there is some $p \in (\Phi_{\psi'} - \Psi)$ then, as above, we have $K_i(p = 1/2) \in V$ and $p = 1/2 \in W$, and, thus, $\psi' = 1/2 \in W$. Therefore, $\psi = 1 \in W$. If there is no primitive proposition $p \in (\Phi_{\psi'} - \Psi)$, then $(M^c, s_V) \models K_i(\psi' \neq 1/2) = 1$, and thus $(\psi' \neq 1/2) = 1 \in V/K_i$. It follows that if ψ'' is the formula that results from removing the disjunct $\psi' = 1/2$ from ψ, then $\psi'' \in V/K_i$. The result now follows from the induction hypothesis. Thus, we can assume without loss of generality that every disjunct of ψ has the form $\psi' = 0$ or $\psi' = 1$. If there is some disjunct $\psi' = k$, $k \in \{0, 1\}$, that mentions a primitive proposition p such that $p \notin \Psi$, then it is easy to check that $K_i(\psi' = 1/2) \in V$. Thus, $(\psi' = 1/2) = 1 \in V/K_i$. Again, it follows that if ψ'' is the formula that results from removing the disjunct $\psi' = k$ from ψ, then $\psi'' \in V/K_i$ and, again, the result follows from the induction hypothesis. Thus, we can assume that $\psi \in \mathcal{L}_n^{K,\hookrightarrow}(\Psi)$. But then $\psi \in V/K_i$, by assumption. ∎

Proposition C.7: *Suppose* $\Psi_1 \subseteq \Psi_2$, $s_V \in S^c_{\Psi_2}$, $s_W = \rho^c_{\Psi_2,\Psi_1}(s_V)$, $\mathcal{K}_i^c(s_V) \subseteq S^c_{\Psi_3}$, *and* $\mathcal{K}_i^c(s_W) \subseteq S^c_{\Psi_4}$. *Then* $\Psi_4 = \Psi_1 \cap \Psi_3$ *and* $\rho^c_{\Psi_3,\Psi_4}(\mathcal{K}_i^c(s_V)) = \mathcal{K}_i^c(s_W)$.

Proof: By the definition of projection, $\Psi_4 \subseteq \Psi_1$. To show that $\Psi_4 \subseteq \Psi_3$, suppose that $p \in \Psi_4$. Since

$\mathcal{K}_i^c(s_W) \subseteq S^c_{\Psi_4}$, $(M^c, s_W) \models K_i(p \vee \neg p)$. Since $\Psi_4 \subseteq \Psi_1$, by Lemma C.1, $(M^c, s_V) \models K_i(p \vee \neg p)$. Thus, $(M^c, s_{V'}) \models p \vee \neg p$ for all $s_{V'} \in \mathcal{K}_i^c(s_V)$. Therefore, $p \in \Psi_3$. Thus, $\Psi_4 \subseteq \Psi_1 \cap \Psi_3$. For the opposite containment, if $p \in \Psi_1 \cap \Psi_3$, then $(M^c, s_V) \models K_i(p \vee \neg p)$. By definition of projection, since $p \in \Psi_1$ and $s_W = \rho^c_{\Psi_2, \Psi_1}(s_V)$, $(M^c, s_W) \models K_i(p \vee \neg p)$. Thus, $p \in \Psi_4$. It follows that $\Psi_4 = \Psi_1 \cap \Psi_3$.

To show that $\rho^c_{\Psi_3,\Psi_4}(\mathcal{K}_i^c(s_V)) = \mathcal{K}_i^c(s_W)$, we first prove that $\rho^c_{\Psi_3,\Psi_4}(\mathcal{K}_i^c(s_V)) \supseteq \mathcal{K}_i^c(s_W)$. Suppose that $s_{W'} \in \mathcal{K}_i^c(s_W)$. We construct V' such that $\rho^c_{\Psi_3,\Psi_4}(s_{V'}) = s_{W'}$ and $s_{V'} \in \mathcal{K}_i^c(s_V)$. We claim that $V/K_i \cup \{\varphi = 1 : \varphi = 1 \in W', \varphi \text{ is implication-free}\}$ is $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. For suppose not. Then there exist formulas $\varphi_1, ..., \varphi_m, \varphi'_1, ..., \varphi'_k$ such that $\varphi_j = 1 \in V/K_i$ for $j \in \{1, ..., m\}$, φ'_j is implication-free and $\varphi'_j = 1 \in W'$ for $j \in \{1, ..., k\}$, and $\{\varphi_1 = 1, ..., \varphi_m = 1, \varphi'_1 = 1, ..., \varphi'_k = 1\}$ is not $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. Let $\psi = \varphi'_1 \wedge ... \wedge \varphi'_k$. Then $\psi = 1 \in W'$, so $\psi = 1/2 \notin W$. Thus, by axiom B1 and Lemma B.4(3), $(K_i \neg\psi) = 1/2 \notin W$. By definition of $\mathcal{K}_i^c$, $(K_i \neg\psi) = 1 \notin W$, for otherwise $\neg\psi = 1 \in W'$. By axiom P11, it follows that $(K_i \neg\psi) = 0 \in W$. Thus, by Lemma B.4(7), $(\neg K_i \neg\psi) = 1 \in W$, so $(\neg K_i \neg\psi) = 1 \in V$. Thus, there must be some maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent set $V'' \supseteq V/K_i$ such that $(\neg\psi) = 1 \notin V''$. Since $s_{V''} \in \mathcal{K}_i^c(s_V) \subseteq S^c_{\Psi_3}$, $\Psi_4 \subseteq \Psi_3$, $s_{W'} \in S_{\Psi_4}$, and $(\neg\psi) = 1/2 \notin W'$, we have $(\neg\psi) = 1/2 \notin V''$. So, by axiom P11, $(\neg\psi) = 0 \in V''$. Thus, $\psi = 1 \in V''$ and $\varphi'_1 = 1, ..., \varphi'_k = 1 \in V''$, which is a contradiction since $V'' \supseteq V/K_i$ and V'' is $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. Let V' be an $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent set containing $V/K_i \cup \{\varphi = 1 : \varphi = 1 \in W', \varphi \text{ is implication-free}\}$. By construction, $s_{V'} \in \mathcal{K}_i^c(s_V)$. Moreover, since $(\neg\varphi) = 1 \rightleftharpoons \varphi = 0$ is valid in $\mathcal{H}_n$ and $s_{W'} \in S^c_{\Psi_4}$, $\rho^c_{\Psi_3,\Psi_4}(s_{V'})$ agrees with $s_{W'}$ on all formulas of the form $\varphi = k$ for $k \in \{0, 1/2, 1\}$ and φ implication-free. Therefore, it is easy to show using Prop$'$ that they agree on all simple formulas. Thus, by the proof of Lemma C.4, they agree on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi_4)$. By uniqueness of ρ^c, it follows that $\rho^c_{\Psi_3,\Psi_4}(s_{V'}) = s_{W'}$, as desired.

The proof of the other direction $\rho^c_{\Psi_3,\Psi_4}(\mathcal{K}_i^c(s_V)) \subseteq \mathcal{K}_i^c(s_W)$ is similar. Suppose that $s_{V'} \in \mathcal{K}_i^c(s_V)$ and $s_{V''} = \rho^c_{\Psi_3,\Psi_4}(s_{V'})$. We need to prove that $s_{V''} \in \mathcal{K}_i^c(S_W)$. We claim that $W/K_i \cup \{\varphi = 1 : \varphi = 1 \in V'', \varphi \text{ is implication-free}\}$ is $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. For suppose not. Then there exist formulas

$\varphi_1, ..., \varphi_m, \varphi'_1, ..., \varphi'_k$ such that $\varphi_j = 1 \in W/K_i$ for $j \in \{1, ..., m\}$, φ'_j is implication-free and $\varphi'_j = 1 \in V''$ for $j \in \{1, ..., k\}$, and $\{\varphi_1 = 1, ..., \varphi_m = 1, \varphi'_1 = 1, ..., \varphi'_k = 1\}$ is not $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. Let $\psi = \varphi'_1 \wedge ... \wedge \varphi'_k$. Then $\psi = 1 \in V''$, so $\psi = 1/2 \notin V$. Thus, by axiom B1 and Lemma B.4(3), $(K_i \neg\psi) = 1/2 \notin V$. Using Lemmas C.1 and C.2, it is easy to show that $\psi = 1 \in V''$ implies $\psi = 1 \in V'$, so, by definition of $\mathcal{K}_i^c$, $(K_i \neg\psi) = 1 \notin V$, for otherwise $\neg\psi = 1 \in V'$. By axiom P11, it follows that $(K_i \neg\psi) = 0 \in V$. Thus, by Lemma B.4(7), $(\neg K_i \neg\psi) = 1 \in V$. But as $s_{V''} \in S^c_{\Psi_4}$, $\psi = 1 \in V''$ implies $\Phi_\psi \in S^c_{\Psi_4}$. Since $\Psi_4 \subseteq \Psi_1$, it follows that $\psi = 1/2 \notin W$, thus, by definition of projection, we have $(\neg K_i \neg\psi) = 1 \in W$. Thus, there must be some maximal $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent set $W' \supseteq W/K_i$ such that $(\neg\psi) = 1 \notin W'$. Since $s_{W'} \in \mathcal{K}_i^c(s_W) \subseteq S^c_{\Psi_4}$, we have $(\neg\psi) = 1/2 \notin W'$. So by P11 it follows that $(\neg\psi) = 0 \in W'$, so $\psi = 1 \in W'$. Thus, $\varphi'_1 = 1, ..., \varphi'_k = 1 \in W'$, which is a contradiction since $W' \supseteq W/K_i$ and W' is $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. Let W' be an $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent set containing $W/K_i \cup \{\varphi = 1 : \varphi = 1 \in V'', \varphi \text{ is implication-free}\}$. By construction, $s_{W'} \in \mathcal{K}_i^c(s_W)$. Moreover, since $(\neg\varphi) = 1 \rightleftharpoons \varphi = 0$ is valid in $\mathcal{H}_n$ and $s_{W'}, s_{V''} \in S^c_{\Psi_4}$, $s_{V''}$ agrees with $s_{W'}$ on all formulas of the form $\varphi = k$ for $k \in \{0, 1/2, 1\}$ and φ implication-free. Therefore, it is easy to show using Prop$'$ that they agree on all simple formulas. Thus, by the proof of Lemma C.4, they agree on all formulas in $\mathcal{L}_n^{K,\hookrightarrow}(\Psi_4)$. By uniqueness of ρ^c, it follows that $s_{V''} = s_{W'} \in \mathcal{K}_i^c(s_W)$, as desired. ∎

The following result is immediate from Proposition C.7.

Corollary C.8: *Projection preserves knowledge and ignorance in M^c.*

Since projections preserve knowledge and ignorance in M^c, it follows that $M^c \in \mathcal{H}_n(\Phi)$. This finishes the proof for the case $\mathcal{C} = \emptyset$. If $T' \in \mathcal{C}$, we must show that M^c satisfies generalized reflexivity. Given $s_V \in S^c_{\Psi_1}$, by confinedness, $\mathcal{K}_i^c(s_V) \subseteq S^c_{\Psi_2}$ for some $\Psi_2 \subseteq \Psi_1$. It clearly suffices to show that $s_W = \rho^c_{\Psi_1,\Psi_2}(s_V) \in \mathcal{K}_i^c(s_V)$. By Lemma C.6, to prove this, it suffices to show that $V/K_i \cap \mathcal{L}_n^{K,\hookrightarrow}(\Psi_2) \subseteq W$. If $\varphi = 1 \in V/K_i \cap \mathcal{L}_n^{K,\hookrightarrow}(\Psi_2)$, then $K_i\varphi = 1 \in V$. Note that Prop$'$ implies that $\mathrm{AX}_n^{K,\hookrightarrow} \vdash \varphi \hookrightarrow \varphi = 1$, which by Gen, K$'$, Prop$'$, and MP$'$ implies that $\mathrm{AX}_n^{K,\hookrightarrow} \vdash (K_i\varphi \hookrightarrow K_i(\varphi = 1)) = 1$. Therefore, as $(K_i\varphi) = 1 \in V$, by Lemma B.4(8,10), $(K_i(\varphi = 1)) = 1 \in V$. By Prop$'$ and Lemma B.4(3), it follows that $K_i(\varphi = 1) \in V$. By T$'$ and Lemma B.4(3), it follows that $\varphi = 1 \vee \bigvee_{\{p:p\in\Phi_\varphi\}} K_i(p = 1/2) \in V$. But since $\varphi \in \mathcal{L}_n^{K,\hookrightarrow}(\Psi_2)$, we must have $\Phi_\varphi \subseteq \Psi_2$. Moreover, since $\mathcal{K}_i^c(s_V) \subseteq S^c_{\Psi_2}$, we must have $\neg K_i(p = 1/2) \in V$ for all $p \in \Psi_2$. Thus, it easily follows that $\varphi = 1 \in V$. Finally, since $s_W = \rho^c_{\Psi_1,\Psi_2}(s_V)$, we must have $\varphi = 1 \in W$, as desired.

Now suppose $4' \in \mathcal{C}$. We want to show that M^c satisfies part (a) of stationarity. Suppose that $s_W \in \mathcal{K}_i^c(s_V)$ and $s_X \in \mathcal{K}_i^c(S_W)$. We must show that $s_X \in \mathcal{K}_i^c(S_V)$. If $(K_i\varphi) = 1 \in V$ then, by axioms Prop$'$ and $4'$ and Lemma B.4(3), $(K_iK_i\varphi) = 1 \in V$. This implies that $(K_i\varphi) = 1 \in W$, which implies that $\varphi = 1 \in X$. Thus, $V/K_i \subseteq X$ and $s_X \in \mathcal{K}_i^c(s_V)$, as desired.

Finally, suppose that $5' \in \mathcal{C}$. We want to show that M^c satisfies part (b) of stationarity. Suppose that $s_W \in \mathcal{K}_i^c(s_V)$ and $s_X \in \mathcal{K}_i^c(s_V)$. We must show that $s_X \in \mathcal{K}_i^c(s_W)$. Suppose by way of contradiction that $s_X \notin \mathcal{K}_i^c(s_W)$, then there exists $\varphi = 1 \in W/K_i$ such that $\varphi = 1 \notin X$. $\varphi = 1 \in W/K_i$ implies that $(K_i\varphi) = 1 \in W$, so by Prop$'$, B1, and Lemma B.4(3) it follows that $\varphi = 1/2 \notin W$. By Conf2 it can be easily shown that $\varphi = 1/2 \notin X$, so by Prop$'$, we get that $\varphi = 0 \in X$. As in the proof of the case $T' \in \mathcal{C}$, if $(K_i\varphi) = 1 \in W$, then $(K_i(\varphi = 1)) = 1 \in W$. Then it follows that $(K_i\neg K_i(\varphi = 1)) = 1 \notin V$, for otherwise since $s_W \in \mathcal{K}_i^c(s_V)$ we get that $(\neg K_i(\varphi = 1)) = 1 \in W$. Since $(K_i\neg K_i(\varphi = 1)) \in D_2$, it is easy to show that $(\neg K_i\neg K_i(\varphi = 1)) = 1 \in V$. By Prop$'$ and $5'$ and Lemma B.4(8,10), it follows that $(K_i(\varphi = 1) \vee K_i((\varphi = 1) = 1/2)) = 1 \in V$. Then using Lemma B.4(5,6,7), it easily follows that either $(K_i(\varphi = 1)) = 1 \in V$ or $(K_i((\varphi = 1) = 1/2)) = 1 \in V$. Then, either $(\varphi = 1) = 1 \in X$ or $((\varphi = 1) = 1/2) = 1 \in X$, but this is a contradiction since $\varphi = 0 \in X$ and X is $\mathrm{AX}_n^{K,\hookrightarrow}$-consistent. ∎